Julien-Daniel Guelfi
Véronique Gaillac / Roland Dardennes

PSYCHOPATHOLOGIE QUANTITATIVE



MASSON

PSYCHOPATHOLOGIE QUANTITATIVE

Sous la direction de J.-D. Guelfi, V. Gaillac, R. Dardennes

La psychopathologie quantitative est une application de la psychométrie aux données de la pathologie mentale. Elle a vu son essor avec le développement des échelles d'évaluation dans les années 60, des critères diagnostiques dans les années 70, puis avec la généralisation de l'emploi de ces derniers par l'*American Psychiatric Association* dans sa classification des troubles mentaux (DSM-III, 1980).

Cet ouvrage offre un panorama des outils disponibles pour la recherche clinique et un aperçu de leur utilisation dans la pratique. Sont ainsi abordés les critères diagnostiques et leur intérêt dans l'approche polydiagnostique, les échelles d'évaluation et leur utilisation dans les essais cliniques, les tests mentaux et les questionnaires pour l'exploration de certaines conduites pathologiques ou l'évolution sous traitement de troubles comportementaux ou cognitifs. La majorité de ces outils concernent des signes, des symptômes ou des conduites pathologiques. D'autres instruments évaluent des concepts comme la qualité de vie, le sentiment de contrôle personnel, l'anhédonie ou encore l'adaptation sociale. Deux chapitres sont consacrés aux bonnes pratiques statistiques des données recueillies.

Cet ouvrage collectif est destiné aux psychiatres, aux étudiants en psychiatrie ainsi qu'à tous les professionnels de la santé mentale, aux psychologues, travailleurs sociaux, épidémiologistes et chercheurs qui doivent interpréter et traiter des données en psychopathologie.

Julien-Daniel GUELFI est professeur de psychiatrie à l'université Paris V, UFR Cochin-Port-Royal-Sainte-Anne.

Véronique GAILLAC et Roland DARDENNES sont praticiens hospitaliers à l'hôpital Sainte-Anne de Paris (clinique des maladies mentales et de l'encéphale).

COUVERTURE ANNIE CIVARD / ILLUSTRATION MARTIN MELKONIAN



ISBN : 2-225-**84871**-8

CHEZ LE MÊME ÉDITEUR

Autres ouvrages :

CIM 10/ICD 10. CLASSIFICATION INTERNATIONALE DES MALADIES. 10e révision. Chapitre V (F). Troubles mentaux et troubles du comportement. Descriptions cliniques et directives pour le diagnostic. Organisation mondiale de la santé. Traduction de l'anglais par C. B. PULL. 1993, 2e tirage, 336 pages.

CIM 10. CLASSIFICATION INTERNATIONALE DES TROUBLES MENTAUX ET DES TROUBLES DU COMPORTEMENT. Critères diagnostiques pour la recherche. Organisation mondiale de la santé. Traduction de l'anglais par C. B. PULL. 1994, 248 pages.

LES THÉRAPIES COMPORTEMENTALES ET COGNITIVES, par J. COTTRAUX. *Collection Médecine et Psychothérapie*. 1995, 2e édition revue et augmentée, 328 pages.

BASES NEUROLOGIQUES DES COMPORTEMENTS, par M. HABIB. Illustrations de J.-F. LANDOU et D. ANSIDEI. 1993, 2e édition, 272 pages, 160 figures.

PSYCHOPATHOLOGIE COGNITIVE, sous la direction de G. LOAS, P. BOYER, B. SAMUEL-LAJEUNESSE. 1991, 184 pages, 15 figures, 4 tableaux.

LES CONDUITES ALIMENTAIRES, sous la direction de B. SAMUEL-LAJEUNESSE, C. FOULON. 1994, 280 pages.

PSYCHOPATHOLOGIE QUANTITATIVE

sous la direction de

J. D. GUELFI V. GAILLAC R. DARDENNES

avec la collaboration de

I. BICHINDARITZ, M. P. BOUVARD, P. BOYER, Ph. CIALDELLA, O. CHAMBON, S. CONSOLI, S. CRIQUILLION-DOUBLET, A. DAZORD, S. DIVAC, S. DOLLFUS, T. DUPIN-SPRIET, B. FALISSARD, J. FERMANIAN, I. GILLET, L. HUGONOT-DIENER, M. HUTEAU, R. JOUVENT, P. KOWAL, V. LECLERC, J.-P. LÉPINE, G. LOAS, Ch. MIRABEL-SARRON, P. PICHOT, J. POITRENAUD, C. B. PULL, M. SIMON, A. SPRIET


MASSON

Paris Milan Barcelone



Ce logo a pour objet d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, tout particulièrement dans le domaine universitaire, le développement massif du « photocopillage ».
Cette pratique qui s'est généralisée, notamment dans les établissements d'enseignement, provoque une baisse brutale des achats de livres, au point que la possibilité même pour les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée.
Nous rappelons donc que la reproduction et la vente sans autorisation, ainsi que le recel, sont passibles de poursuites. Les demandes d'autorisation de photocopier doivent être adressées à l'éditeur ou au Centre français d'exploitation du droit de copie : 3, rue Hautefeuille, 75006 Paris. Tél. : 43 26 95 35.

ISBN 2-225-84871-8

MASSON S.A. 120, bd Saint-Germain, 75280 Paris Cedex 06
MASSON S.p.A. Via Statuto 2/4, 20121, Milano
MASSON S.A. Avenida Principe de Asturias 20, 08012 Barcelona

LISTE DES COLLABORATEURS

BICHINDARITZ I. : docteur en informatique de l'université René-Descartes, service du professeur B. SAMUEL-LAJEUNESSE, hôpital Sainte-Anne, Paris.

BOUVARD M. P. : praticien hospitalier, service du professeur M. C. MOUREN-SIMEONI, hôpital Robert-Debré, Paris.

BOYER P. : chargé de recherche, INSERM, Paris.

CIALDELLA Ph. : ancien assistant, service du docteur ESCOLIER, hôpital Saint-Jean-de-Dieu, Lyon.

CHAMBON O. : ancien assistant, chef de clinique, Lyon.

CONSOLI S. : professeur de psychiatrie, hôpital Broussais, Paris.

CRIQUILLION-DOUBLET S. : praticien hospitalier, service du professeur B. SAMUEL-LAJEUNESSE, hôpital Sainte-Anne, Paris.

DARDENNES R. : praticien hospitalier, service du professeur B. SAMUEL-LAJEUNESSE, hôpital Sainte-Anne, Paris.

DAZORD A. : directeur de recherche, INSERM, Lyon.

DIVAC S. : docteur en psychologie, service du professeur B. SAMUEL-LAJEUNESSE, hôpital Sainte-Anne, Paris.

DOLLFUS S. : professeur de psychiatrie, service du professeur E. ZARIFIAN, Centre hospitalo-universitaire Côte-de-Nacre, Caen.

DUPIN-SPRIET T. : conseil, Paris.

FALISSARD B. : interne des hôpitaux, biostatisticien, Paris.

FERMANIAN J. : professeur, laboratoire de biostatistiques, hôpital Necker, Paris.

GAILLAC V. : praticien hospitalier, service du professeur B. SAMUEL-LAJEUNESSE, hôpital Sainte-Anne, Paris.

GILLET I. : psychologue, éditions du Centre de psychologie appliquée, Paris.

GUELFI J.-D. : professeur de psychiatrie, service du professeur B. SAMUEL-LAJEUNESSE, hôpital Sainte-Anne, Paris.

HUGONOT-DIENER L. : directeur du Centre de prévention, d'études et de recherches sur le vieillissement (CEPEVI), Paris.

HUTEAU M. : professeur de psychologie, CNAM-INETOP (institut national d'études du travail et d'orientation professionnelle), Paris.

JOUVENT R. : directeur de l'URA 1957, CNRS, hôpital de la Salpêtrière, Paris, Professeur Paris VII, Psychopathologie et Biologie.

KOWAL P. : psychiatre, ancien interne des hôpitaux, Paris.

LECLERC V. : psychologue, service du professeur B. SAMUEL-LAJEUNESSE, hôpital Sainte-Anne, Paris.

LEPINE J.-P. : professeur de psychiatrie, service du professeur J. ADES, hôpital Fernand-Widal, Paris.

LOAS G. : praticien hospitalier, service du professeur DELAHOUSSE, Amiens.

MIRABEL-SARRON Ch. : praticien hospitalier, service du professeur B. SAMUEL-LAJEUNESSE, hôpital Sainte-Anne, Paris.

PICHOT P. : professeur de psychiatrie, membre de l'Académie de médecine, Paris.

POITRENAUD J. : directeur de recherche Paris V, Centre de gérontologie de l'association Claude-Bernard, Paris.

PULL C.-B. : professeur de psychiatrie, Centre hospitalier du Luxembourg, Grand Duché du Luxembourg.

SIMON M. : psychologue, éditions du Centre de psychologie appliquée, Paris.

REMERCIEMENTS

Les trois auteurs-coordonnateurs de cet ouvrage collectif tiennent à remercier tous ceux qui ont participé au séminaire annuel de la Clinique des maladies mentales et de l'encéphale (service du professeur B. Samuel-Lajeunesse en 1993-1994) soit en préparant les exposés puis les textes écrits, soit en animant, au fil des séances, les discussions.

Le soutien de l'association « Psychiatrie et Psychobiologie » a aussi été pour nous une aide appréciable.

Nous tenons enfin à remercier tout particulièrement Madame Marie-Hélène Mayengo-Cécile qui a assuré l'ensemble du secrétariat du séminaire avec compétence et bonne humeur.

AVANT-PROPOS

La psychopathologie quantitative est une application de la psychométrie aux données de la pathologie mentale. La psychométrie, « art d'imposer aux opérations de l'esprit la mesure et le nombre » (Galton, 1879), englobe les *tests mentaux* (dont les questionnaires font partie), les *échelles d'évaluation*, et les listes à cocher de critères diagnostiques ou *échelles-critères*.

Le pionnier français de la psychopathologie quantitative fut Pierre Pichot qui a dirigé la Clinique des maladies mentales et de l'encéphale après Jean Delay, de 1969 à 1986. C'est en hommage à son œuvre que fut organisé en 1993-1994, avec le concours de Roland Dardennes et de Véronique Gaillac, le séminaire annuel de la Clinique sur ce thème.

Si les échelles d'évaluation ont été très populaires dans la psychiatrie anglo-saxonne dès le milieu des années 1950, avec, aux États-Unis, des auteurs comme J. R. Wittenborn ou J. E. Overall et, en Angleterre, M. Hamilton, il n'en fut pas de même en France, où la psychopathologie quantitative a été initialement très largement critiquée « au nom de l'impossibilité supposée de traduire en chiffres le caractère ineffable du contact interpersonnel » (P. Pichot) et où elle a été accueillie avec la plus grande « méfiance par les psychiatres pour son origine psychologique » (id).

Par ailleurs, jusqu'en 1969-1970, il était parfois considéré comme de bon ton, dans certains milieux, de se déclarer hostile aux classements, de railler les étiquettes et de se proclamer antinosographiste.

Les temps ont changé. L'assertion de C. Lévi-Strauss selon laquelle tout classement est supérieur au chaos est généralement acceptée et le refus d'évaluation fait de plus en plus figure de position défensive... indéfendable.

Et puisqu'il faut bien se résoudre à évaluer, l'utilisation d'outils standardisés et la manipulation des chiffres sont deux procédés — arbitraires certes — mais, en définitive, peut-être moins dangereux que d'autres.

Le début des années 1970 — avec le développement aux États-Unis des critères diagnostiques — a marqué un nouvel essor de la psychopathologie quantitative et la classification américaine des troubles mentaux publiée en 1980 par l'association américaine de psychiatrie, connue sous le nom de DSM-III (*Diagnostic and Statistical Manual, 3e édition*) a consacré la généralisation de l'emploi du procédé des critères diagnostiques.

Cet ouvrage collectif est destiné aux psychiatres, aux étudiants en psychiatrie ainsi qu'à tous les professionnels de la santé mentale, aux psychologues, travailleurs sociaux, épidémiologistes et chercheurs qui doivent interpréter et traiter des données cliniques.

Notre premier objet de réflexion concerne les critères diagnostiques. Quels sont les différents types de critères utilisables? Et peut-on espérer « formaliser », « opérationnaliser » les procédures intellectuelles qui conduisent un psychiatre à porter un diagnostic en psychiatrie? Puis vient le domaine des échelles d'évaluation en psychopathologie. L'histoire exemplaire du développement de la *Brief Psychiatric Rating Scale* ou BPRS de J. E. Overall et D. R. Gorham a été choisie pour illustrer l'évolution des techniques utilisées en psychopathologie quantitative au cours des trente dernières années. Sont ensuite envisagées les modalités de choix, de construction et d'utilisation des échelles d'évaluation.

Plusieurs exemples de domaines d'application ont été choisis, comme l'intérêt de la démarche *polydiagnostique* par l'utilisation conjointe de différents systèmes de critères diagnostiques au cours de la schizophrénie, ou la description actualisée d'un ensemble symptomatique récemment redécouvert, le syndrome mixte anxieux et dépressif.

Le domaine de certains *tests mentaux* est ensuite abordé : tests cognitifs, batteries informatisées, évolution de la psychométrie de l'intelligence depuis le début du siècle jusqu'à la nouvelle clinique. Comment peut-on étudier de nos jours le vieillissement cognitif normal? L'intelligence artificielle peut-elle aider la recherche clinique?

D'autres problèmes d'actualité concernent la psychométrie de la personnalité. Que penser aujourd'hui de la distinction trait-état? L'heure est-elle venue d'une psychobiologie de la personnalité? Le débat — maintenant ancien — sur la permanence contestée des traits de personnalité est-il définitivement dépassé par le point de vue interactioniste? Y a-t-il un renouveau des questionnaires de personnalité, avec le MMPI-2 par exemple?

D'autres champs d'application importants des questionnaires concernent l'exploration de certaines conduites pathologiques, ou l'évolution sous traitement de symptômes ou de troubles du comportement. Il en est ainsi de l'évaluation des dysfonctionnements lors des approches cognitives, de l'évaluation du résultat en thérapie comportementale ou, d'une façon plus générale, de l'évaluation lors des psychothérapies ou de l'exploration des conduites chez les enfants.

Si la majorité des échelles d'évaluation utilisées en psychiatrie concernent des signes, des symptômes ou des conduites pathologiques voire des ensembles syndromiques, certains instruments, de développement relativement récent, explorent d'autres secteurs, moins clairement délimités, partiellement liés aux symptômes seulement.

Ce sont, par exemple, les instruments évaluant la qualité de vie, les attentes, l'état de satisfaction ou d'insatisfaction ou la qualité de l'adaptation sociale des patients.

Mais ces nouveaux instruments peuvent aussi concerner des caractéristiques psychologiques générales comme la capacité à éprouver du plaisir (ou la disparition de celle-ci, l'anhédonie), l'appétence alimentaire ou la satiété, etc.

Tous ces secteurs sont autant de domaines où la psychopathologie quantitative actuelle se montre particulièrement active.

Deux chapitres sont spécifiquement consacrés aux bonnes pratiques statistiques des données recueillies. Des logiciels d'analyse statistique sont désormais à la portée de quasiment tous les utilisateurs néophytes de micro-ordinateurs. Mais les modalités de classement des sujets ou des variables cliniques, l'adéquation entre une question posée et la technique d'analyse à retenir, les règles à respecter dans l'interprétation des résultats obtenus, sont affaires de spécialistes.

D'autre part, nos données cliniques sont notoirement insuffisantes pour fournir les clés d'une véritable connaissance. Les instruments psychométriques que nous utilisons ne sont que des outils de travail. Il est solidement établi qu'il est préférable de les utiliser plutôt que de s'en passer lorsqu'on fait de la recherche clinique et que l'on désire communiquer avec autrui à ce propos. Mais ce serait une erreur que de considérer que ces outils représentent la connaissance.

Les dernières pages de cet ouvrage confirment cette nécessaire humilité. Plusieurs théorèmes ou paradoxes dans d'autres champs que ceux de la psychiatrie sont là en effet pour nous signifier que l'opérationnalisation a des limites obligées et que la toute-puissance des chiffres appartient au monde des fantasmes.

J. D. GUELFI

TABLE DES MATIÈRES

1

L'OPÉRATIONNALISATION DU DIAGNOSTIC PSYCHIATRIQUE INTÉRÊT ET LIMITES[1]

par C. B. PULL et J. D. GUELFI

DÉFINITIONS ET POSITION DU PROBLÈME

Les appellations « critères diagnostiques » et « critères opérationnels » se réfèrent, en psychiatrie, à un procédé dont les principes et les modalités ont été élaborés au début des années 70 par un groupe de chercheurs de la Washington University à Saint-Louis (Feighner *et al.*, 1972). Ce procédé a pour objectif la définition rigoureuse et explicite de chacune des opérations qui conduisent à un diagnostic psychiatrique. Il s'applique, schématiquement, à trois types de critères :

– aux critères de classification des troubles, c'est-à-dire aux principes à partir desquels les classifications sont élaborées ;

– aux critères de classement des sujets, c'est-à-dire aux règles que suit le clinicien pour vérifier à quel diagnostic répond la pathologie observée dans chaque cas ;

– aux critères d'évaluation du clinicien, c'est-à-dire aux modalités d'investigation utilisées par le clinicien pour mettre en évidence des éléments psychopathologiques.

Réservé initialement au seul domaine de la recherche, ainsi que l'indique l'appellation de « *Diagnostic Criteria for Use in Psychiatric Research* », ou celle de « *Research Diagnostic Criteria* » ou RDC, proposée un peu plus tard par Spitzer, le procédé a, par la suite, été appliqué de façon systématique dans le *Manuel diagnostique et statistique des troubles mentaux* ou DSM (APA, 1980, 1987, 1994), à partir de sa troisième édition. Il constitue également la base de l'une des versions de la 10e révision du chapitre V(F) de la classification internationale des maladies ou CIM-10 (OMS, 1993), à savoir celle qui est proposée pour la recherche, sous l'appellation de « *critères diagnostiques pour la recherche* » ou CDR-10 (OMS, 1994).

En dépit de contestations parfois véhémentes, la nouvelle approche s'est rapidement imposée, auprès des cliniciens comme des chercheurs, au point que, souvent, on en arrive à oublier qu'elle est très récente, et que, malgré des avantages incontestés, elle présente également des inconvénients et des limites.

1. Plusieurs paragraphes de ce chapitre figurent et sont développés dans : PULL C. B., GUELFI J. D. et PULL M. C. – *Les critères diagnostiques en psychiatrie*. Éditions techniques, Encycl. Méd. Chir. Psychiatrie, 37102 C-15, Paris, 1995 (note des éditeurs).

LES CRITÈRES DE CLASSIFICATION

Les critères de classification constituent les éléments de base sur lesquels repose l'édifice entier de la classification. Ces critères peuvent être de deux types différents : pathogénique ou descriptif, et le recours à l'un ou à l'autre va aboutir à deux modèles psychopathologiques fondamentalement différents (Pull, Guelfi, 1995). Le premier est fondé sur la notion de maladie et postule l'existence d'entités naturelles définies avant tout par leur étiologie et leur pathogénie. Le deuxième repose sur la description de syndromes, c'est-à-dire de constellations de signes et de symptômes associés avec une fréquence plus grande que ne le laisserait supposer une distribution au hasard.

D'après le procédé de Saint-Louis, les critères utilisés dans la classification des troubles psychiatriques doivent être athéoriques par rapport à l'étiologie ou la physiopathologie, sauf dans les cas où l'étiologie est connue. L'approche préconisée pour le regroupement des troubles psychiatriques en classes diagnostiques est descriptive, les troubles démunis d'étiologie ou de processus physiopathologique connu étant groupés selon des caractéristiques descriptives communes.

L'*approche descriptive* permet de préciser sans ambiguïté, en d'autres termes d'opérationnaliser, chacun des critères de classification utilisés pour définir un trouble et, éventuellement, pour le délimiter par rapport à un autre trouble. Les critères de classification descriptifs concernent essentiellement la présence de signes et de symptômes caractéristiques, leur durée, leur fréquence, l'ordre de leur survenue par rapport à d'autres manifestations, et éventuellement leur degré de sévérité ou leur retentissement sur le plan psychosocial.

L'approche athéorique présente deux avantages incontestés. En premier lieu, elle permet à des cliniciens d'orientations théoriques différentes de communiquer à l'aide d'un même système de référence. En deuxième lieu, de nombreuses études ont démontré l'efficacité de l'approche descriptive pour l'amélioration de la fidélité inter-juges du diagnostic psychiatrique.

Il convient toutefois de relever que l'approche descriptive présente également des inconvénients et des limites.

Le recours systématique à des critères de classification descriptifs a conduit à un éclatement des anciennes maladies — en l'occurrence principalement des psychoses et des névroses — en une multitude de syndromes. À titre d'exemple, la catégorie « hystérie de conversion » de la CIM-9 a été remplacée, dans la CIM-10, par une douzaine de « troubles dissociatifs ».

Ceci soulève trois types de difficultés. En premier lieu, les nouvelles catégories constituent souvent des constructions quelque peu artificielles, et certaines d'entre elles ne sont observées que rarement, à l'état pur, dans la pratique courante. En deuxième lieu, la validité de la plupart de ces catégories reste encore à établir. Enfin, et surtout, un grand nombre de patients répondent simultanément aux définitions de plusieurs de ces catégories, ce qui oblige le clinicien à faire des diagnostics multiples, ou de recourir à des diagnostics composites tels que « double dépression » ou « trouble panique avec agoraphobie ».

L'élimination de la plupart des règles hiérarchiques pour structurer une classification doit également être considérée comme une prise de position descriptive. Sur le plan théorique, cette position est parfaitement justifiée, dans la mesure où l'on ne dispose pas, le plus souvent, de preuves empiriques permettant d'éliminer un diagnostic au profit d'un autre. Elle a, par ailleurs, permis d'améliorer considérable-

ment la méthodologie des études épidémiologiques concernant la prévalence et la comorbidité des troubles mentaux (Robins et Regier, 1991). Sur le plan pratique, elle soulève toutefois d'importants problèmes, étant donné que le clinicien devra, lui, faire un choix, et décider à quel trouble il convient d'accorder la priorité, que ce soit pour l'encodage à des fins statistiques, le remboursement par des Caisses de maladie, l'inclusion dans un protocole de recherche, ou la prise en charge et le traitement.

LES CRITÈRES DE CLASSEMENT

D'après le procédé de Saint-Louis, chaque trouble doit être défini par des critères de classement explicites, précis, rigoureux et univoques, réunis sous la forme d'un algorithme.

Les avantages essentiels des critères de classement explicites concernent, d'une part la fidélité inter-juges du diagnostic — nettement plus élevée avec les systèmes à critères qu'avec les systèmes à simple glossaire — et de l'autre, l'homogénéité des échantillons cliniques, nettement plus élevée pour des syndromes rigoureusement délimités à l'aide de critères très sélectifs que pour des troubles plus globaux dont le diagnostic repose sur un jugement global.

Le recours à des critères de classement explicites présente toutefois également des problèmes et des limites.

En premier lieu, le procédé implique que seuls des éléments clairement définissables soient retenus comme critères diagnostiques. Ceci revient à favoriser des critères grossiers, apparents, concernant avant tout des comportements aisément objectivables, mais qui n'ont pas forcément la meilleure pertinence pour le diagnostic.

En deuxième lieu, et surtout, chaque critère individuel doit être quantifié. Cette quantification est opérée de deux façons différentes. Certains critères sont monothétiques, c'est-à-dire qu'ils sont obligatoirement présents. D'autres sont polythétiques, c'est-à-dire réunis dans une liste d'éléments, dont un certain nombre — mais aucun en particulier — est requis pour le diagnostic. Le principe même du procédé oblige ainsi à prendre des décisions catégoriques concernant d'une part la nature, le nombre, la fréquence de survenue, et la durée des signes et des symptômes retenus pour un diagnostic, et de l'autre la valeur ou le poids diagnostique qu'il convient d'attribuer à chaque critère individuel.

Dans l'ensemble, le procédé des critères implique le recours à une « grammaire diagnostique » (Robins, 1889) rigoureuse, dans laquelle toute formulation imprécise est considérée comme une « faute ». À titre d'exemple, des formulations telles que « souvent », « persistant », « la plupart du temps », « aigu », ou « plusieurs », doivent être considérées comme des imperfections qu'il s'agira d'éliminer.

À chacune de ses étapes, le procédé oblige ainsi à prendre des décisions, même quand il n'existe pas, dans l'état actuel des connaissances, d'arguments empiriques pour choisir entre plusieurs alternatives. Ceci soulève la question de la validité des critères diagnostiques explicites actuels.

Depuis les travaux initiaux de l'école de St-Louis, on admet qu'une certaine validité est assurée pour autant que l'on construise des catégories qui correspondent à un tableau clinique précis, à une évolution prévisible et, éventuellement, à une histoire familiale particulière. Or, à l'heure actuelle, même ces conditions, pourtant relativement faciles à vérifier, sont loin d'être remplies dans de nombreux cas. Pour la

plupart des catégories, la validité des critères servant à les délimiter se résume à une validité de «face», c'est-à-dire qu'ils correspondent à un consensus d'experts, ayant eu à leur disposition les connaissances scientifiques de l'époque, et, éventuellement à une validité «discriminante», c'est-à-dire qu'ils permettent de délimiter les catégories diagnostiques les unes par rapport aux autres.

Les véritables études de validation restent encore trop rares à ce jour. Les critères symptomatiques — qui sont le plus souvent les seuls dont nous disposions pour l'instant — ont de toutes les manières leurs limites et il est vraisemblable que de véritables progrès nosographiques ne seront possibles que lorsqu'on disposera conjointement de critères concernant l'évolution, la réactivité thérapeutique, la dimension psychodynamique, l'imagerie cérébrale, la génétique et les corrélations biologiques. Le *Source Book* (association américaine de psychiatrie) qui accompagne le DSM-IV, se propose justement de faire le point sur l'état actuel de cette question.

L'obtention d'échantillons homogènes présente également des limites. En élaborant des critères très sélectifs pour des troubles très nettement délimités, on augmente certes les chances de créer des sous-groupes de patients suffisamment homogènes, mais on est obligé de rejeter dans des catégories atypiques ou résiduelles de très nombreux patients. Si l'on se montre moins restrictif, pour pouvoir classer le maximum de patients dans des formes cliniques non résiduelles, on augmente alors les risques de ne pas pouvoir atteindre l'objectif d'homogénéité des groupes. En d'autres termes, l'attitude à adopter lors du choix des critères n'est pas la même dans le domaine de la recherche, où l'on souhaite obtenir des échantillons certes homogènes — mais forcément particuliers — et dans le domaine de la pratique clinique courante, où l'on souhaite classer correctement la quasi-totalité des patients. La publication de la CIM-10, sous plusieurs versions différentes, adaptées chacune à des objectifs différents, constitue sur ce point une initiative particulièrement intéressante.

Enfin, les critères polythétiques soulèvent de nombreux problèmes théoriques, en partie non résolus jusqu'ici. En effet, le recours à des listes de signes et de symptômes suppose que l'hypothèse du *scaling* soit vérifiée. En particulier, les manifestations en question doivent être transférables, c'est-à-dire que chaque signe ou symptôme doit pouvoir être remplacé par n'importe lequel des autres, et la note-seuil doit permettre de différencier entre trouble authentique et simple agencement de symptômes, ne présentant pas de signification pathologique. Or, il n'existe quasiment pas d'études empiriques ayant apporté de réelles justifications aux seuils diagnostiques choisis, lesquels restent ainsi, dans une certaine mesure, arbitraires.

LES CRITÈRES D'ÉVALUATION

Les critères diagnostiques explicites deviennent opérationnels à partir du moment où chacune des opérations impliquées dans leur évaluation a été définie et précisée. Ceci est réalisé au moyen d'entretiens structurés, dans lesquels chaque critère est traduit en une ou plusieurs questions devant permettre de procéder à une évaluation rigoureuse des éléments qui le composent. Les questions ont pour but de mettre en évidence chacun des éléments pathologiques recherchés, de préciser si ces éléments sont significatifs sur le plan clinique, de noter leur durée et le moment de leur survenue, d'établir s'ils constituent une déviation significative par rapport à un état antérieur prémorbide ou s'ils ont toujours été présents, et de vérifier si leur présence

est effectivement en rapport avec un trouble psychiatrique spécifique et non pas avec une maladie somatique ou la prise d'une substance psychoactive.

De nombreux entretiens structurés ont été développés au cours des dernières années.

À l'initiative du NIMH, Robins *et al.* (1989) ont développé un entretien entièrement structuré destiné à l'évaluation des troubles définis dans le DSM-III : le *Diagnostic Interview Schedule* (DIS), dont l'utilisation est recommandée dans le cadre des études épidémiologiques. Le DIS a été adapté au DSM-III-R et une nouvelle version est actuellement en voie d'élaboration pour le DSM-IV.

Pour le recueil des données requises dans le système des RDC, Spitzer *et al.* (1985) ont proposé le *Schedule for Affective Disorders and Schizophrenia* (SADS), dont il existe trois versions complémentaires : le SADS *regular*, le SADS-L (*lifetime*) et le SADS-C (*change*). Le SADS *regular* permet de recueillir des informations détaillées concernant l'épisode actuel et des informations plus sommaires concernant les antécédents personnels. Le SADS-L est consacré aux évaluations diagnostiques « longitudinales » et le SADS-C à l'évaluation du changement.

Le *Structured Clinical Interview for DSM-III-R* (SCID) a été développé par Spitzer *et al.* (1990) pour l'évaluation des troubles mentaux définis dans le DSM-III-R.

L'OMS a développé trois entretiens structurés. Ces instruments permettent d'évaluer non seulement les CDR-10, mais également les critères du DSM-III-R, et ils seront adaptés pour permettre l'évaluation des critères du DSM-IV.

Le *Composite International Diagnostic Interview* ou CIDI (OMS, 1993, Robins et Wing, 1988) est un instrument diagnostique dont l'utilisation n'est pas réservée à des cliniciens. Il peut également être employé par des non-cliniciens, ne possédant aucune expérience dans le domaine de la psychiatrie. L'interviewer doit se limiter à poser une série de questions préétablies et à noter les réponses du sujet. Dans de nombreux cas, une réponse positive à une question oblige « à préciser » (probe) les conditions dans lesquelles un symptôme s'est manifesté. La construction du CIDI et les conditions de son administration en font un instrument particulièrement intéressant pour les enquêtes épidémiologiques, mais il peut également être utilisé dans d'autres études, ainsi que dans la pratique clinique courante.

Les *Schedules for Clinical Assessment in Neuropsychiatry*, ou SCAN [10, 20], constituent un ensemble d'instruments diagnostiques développés par J. Wing en collaboration avec l'OMS (1990, 1994). Le SCAN comprend en particulier la 10[e] version du PSE (*Present State Examination*), un volumineux glossaire, dans lequel chaque signe et chaque symptôme évalués dans le PSE sont définis, et un programme de diagnostic par ordinateur CATEGO. À l'inverse du CIDI, le SCAN doit être utilisé par des cliniciens expérimentés. Le clinicien est, certes, invité à poser une série de questions préétablies, mais il est autorisé à en poser d'autres, et il doit faire intervenir son propre jugement dans la cotation. Instrument diagnostique très précis, le SCAN est destiné en priorité à la recherche clinique et à l'enseignement.

L'*International Personality Disorder Examination*, ou IPDE (Loranger, Sartorius *et al.*, 1994, OMS, 1994) est un instrument diagnostique réservé à l'évaluation des troubles de la personnalité. Pour chacun des critères diagnostiques proposés, l'IPDE propose plusieurs questions devant permettre à l'évaluateur de conclure à la présence ou à l'absence de comportements ou de traits caractérisant le fonctionnement au long cours du sujet depuis le début de l'âge adulte.

L'intérêt essentiel des instruments structurés réside dans le fait qu'ils permettent de procéder à une évaluation systématique et approfondie de chaque critère diagnostique, selon des règles, des définitions et des instructions précises et éprouvées, et donc en principe adaptées et correctes. De cette façon, ils permettent à l'évaluateur d'éviter de porter des jugements personnels idiosyncrasiques.

L'avantage pratique de ces instruments concerne une fois de plus la fidélité interjuges du diagnostic, nettement plus élevée avec les entretiens structurés qu'avec le simple examen clinique.

L'utilisation des entretiens structurés présente toutefois également des inconvénients et des limites.

En premier lieu, la traduction des critères en questions susceptibles d'engendrer des réponses valides, c'est-à-dire permettant d'évaluer correctement un vécu ou un comportement, présente, en elle-même, des limites. Certains critères peuvent être facilement traduits en questions. Il en est ainsi, par exemple, des critères concernant l'utilisation de substances psychoactives. D'autres critères sont plus difficiles à évaluer à l'aide de questions préétablies. Tel est le cas, par exemple, des critères concernant la vie intime du sujet, son vécu personnel, ou son identité. On doit se demander, en particulier, s'il est possible d'évaluer correctement, à l'aide de quelques questions, un critère tel que «perturbation marquée et persistante de l'identité, caractérisée par une incertitude dans le choix des valeurs». Les questions «*Avez-vous du mal à savoir ce qui est important dans la vie?*» ou «*Avez-vous du mal à distinguer ce qui est bon de ce qui est mauvais?*», proposée dans l'IPDE, paraissent, *a priori*, trop superficielles pour évaluer ce critère.

En deuxième lieu, il convient de préciser que les entretiens structurés permettent, certes, aux évaluateurs d'éviter, dans une large mesure, de porter des jugements subjectifs, mais ils ne sauraient pas éviter bien évidemment que les sujets évalués répondent, eux, de façon subjective. Le sujet évalué doit, forcément, faire intervenir son propre jugement, et ses réponses vont dépendre, entre autres, de la façon dont il comprend ou interprète les questions, de ses capacités d'*insight*, et de son désir de répondre.

Les réponses au sujet vont également être influencées par ses capacités de remémoration, et de ses aptitudes à savoir préciser quand un symptôme est survenu pour la première fois, s'il est survenu en même temps que certains autres symptômes, s'il a persisté aussi longtemps que ces derniers, ou encore si sa survenue a précédé ou suivi certains symptômes déterminés. Il conviendra, de toute façon, de tenir compte de l'âge des sujets ainsi que de leurs capacités intellectuelles.

Enfin, il convient de préciser que l'utilisation correcte des entretiens structurés demande beaucoup de temps, allant de deux heures en moyenne pour le CIDI à trois ou quatre heures pour l'IPDE. Par ailleurs, elle présuppose une formation approfondie, dans des centres spécialisés, pendant au moins une semaine, suivie d'une période, variable selon les instruments, pendant laquelle les utilisateurs doivent s'entraîner sous le contrôle d'un formateur expérimenté.

CONCLUSION

De ce qui précède, il ressort qu'il est actuellement possible, dans une large mesure, d'opérationnaliser les trois types de critères impliqués dans le diagnostic

clinique. Par ailleurs, des efforts considérables sont faits pour perfectionner le procédé dans chacun de ses registres.

Pour ce qui est des critères de classification, le DSM-IV comportera d'une part une réorganisation de l'ensemble des troubles organiques, et de l'autre un nombre important de nouveaux syndromes.

Pour ce qui est des critères de classement, la CIM-10 tente de résoudre certains problèmes en proposant une demi-douzaine de versions complémentaires, parfaitement compatibles entre elles, mais adaptées à des utilisateurs différents.

Enfin, pour ce qui est des critères d'évaluation, des versions révisées des entretiens structurés sont présentées avec une rapidité telle que leurs auteurs ont ressenti la nécessité d'éditer des newsletters pour tenir les utilisateurs au courant de l'évolution de leurs travaux.

Pour l'instant, le processus d'opérationnalisation est ainsi en pleine expansion et ne semble guère avoir de limites. Son intérêt essentiel, incontesté, concerne la fidélité inter-juges du diagnosic. Reste à savoir s'il permettra de contribuer à faire des progrès dans d'autres domaines, en particulier dans celui de la validité des catégories diagnostiques et des critères utilisés dans leur définition.

BIBLIOGRAPHIE

Association américaine de psychiatrie — *Manuel diagnostique et statistique des troubles mentaux.* Troisième édition, DSM-III 1980. Traduction de l'anglais coordonnée par GUELFI J. D., Masson, Paris, 1983.

Association américaine de psychiatrie — *Manuel diagnostique et statistique des troubles mentaux.* Troisième édition révisée, DSM-III R, 1987. Traduction de l'anglais coordonnée par GUELFI J. D., Masson, Paris, 1989.

Association américaine de psychiatrie — *Manuel diagnostique et statistique des troubles mentaux.* Quatrième édition, DSM-IV, 1994. Traduction de l'anglais coordonnée par GUELFI J. D., Masson, Paris (en préparation).

Association américaine de psychiatrie — *DSM-IV Source Book.* American Psychiatric Association, Washington, Vol. 1, 1994.

FEIGHNER J. P., ROBINS E., GUZE S. B., WOODRUFF R. A., WINOKUR G., MUNOZ R. — Diagnostic Criteria for Use in Psychiatric Research. *Arch. Gen. Psychiatry*, *26*, 57-63, 1972.

LORANGER W. A., SARTORIUS N., *et al.* — The International Personality Disorder Examination. The World Health Organization/Alcohol, Drug Abuse, and Mental Health Administration International Pilot Study of Personality Disorders. *Arch. Gen. Psychiatry*, *51*, 215-224, 1994.

Organisation mondiale de la santé — *Troubles mentaux : glossaire et guide de classification en concordance avec la neuvième révision de la classification internationale des maladies.* Organisation mondiale de la santé, Genève, 1979.

Organisation mondiale de la santé — *Classification internationale des troubles mentaux et des troubles du comportement. Descriptions cliniques et directives pour le diagnostic.* Traduction de l'anglais coordonnée par PULL C. B. OMS, Genève et Masson, Paris, 1993.

Organisation mondiale de la santé — *Classification internationale des troubles mentaux et des troubles du comportement. Critères diagnostiques pour la recherche.* Traduction de l'anglais coordonnée par PULL C. B. OMS, Genève et Masson, Paris, 1994.

Organisation mondiale de la santé — *SCAN : Schedules for Clinical Assessment in Neuropsychiatry*. Traduction de l'anglais coordonnée par PULL C. B. Organisation mondiale de la santé, Genève, 1994.

Organisation mondiale de la santé — *The International Personality Disorder Examination (IPDE), version 1.1*. Traduction de l'anglais coordonnée par PULL C. B. Organisation mondiale de la santé, Genève, 1994.

Organisation mondiale de la santé — *Composite International Diagnostic Interview (CIDI), core version 1.1*. Traduction de l'anglais coordonnée par PULL C. B. Organisation mondiale de la santé, Genève, 1993.

PULL C. B., GUELFI J. D., BOYER P., PULL M. C. — *Les critères diagnostiques en psychiatrie : historique, état actuel et perspectives d'avenir*. Masson, Paris, 1986.

PULL C. B., GUELFI J. D., PULL M. C. — *Les critères diagnostiques en psychiatrie*. Éditions techniques. Encycl. Méd. Chir. (Paris, France),Psychiatrie, 37-102 C15, 1995.

ROBINS L. N., HELZER J. E., CROUGHAN J., RATCLIFF K. — National Institute of Mental Health Diagnostic Interview Schedule : its History, Characteristics, and Validity. *Arch. Gen. Psychiat.*, *38*, 381-389, 1981.

ROBINS L. N. — Diagnostic Grammar and Assessment : Translating Criteria into Questions. *Psychological Medicine*, *19*, 57-68, 1989.

ROBINS L. N., REGIER D. A. (eds) — *Psychiatric Disorders in America. The Epidemiologic Catchment Area Study*. Free Press, New York, 1991.

ROBINS L. N., WING J. K., *et al.* — The Composite International Diagnostic Interview : an Epidemiologic Instrument Suitable for Use in Conjunction with Different Diagnostic Systems and in Different Cultures. *Arch. Gen. Psychiatry*, *45*, 1069-1077, 1988.

SPITZER R. L., ENDICOTT J., ROBINS E. — Research Diagnostic Criteria (RDC) for a Selected Group of Functional Disorders. New York State Psychiatric Institute, New York, 1977. Traduction de l'anglais par ANSSEAU M., Acta Psychiatr. Belg, 85, 253-324, 1985.

SPITZER R. L., WILLIAMS J. B. W., GIBBON M., FIRST M. B. — *SCID : Structured Clinical Interview for DSM-III-R*. New York State Psychiatric Institute, New York, 1990.

WING J. K., BABOR T., BRUGHA T., *et al.* — SCAN : Schedules for Clinical Assessment in Neuropsychiatry. *Arch. Gen. Psychiatry*, *47*, 589-593, 1990.

2

HISTORIQUE DU DÉVELOPPEMENT D'UNE ÉCHELLE : À PROPOS DE LA BPRS

par P. PICHOT

L'échelle abrégée d'appréciation psychiatrique (*Brief Psychiatric Rating Scale*, ou BPRS), décrite en 1962 par John E. Overall et D. R. Gorham, disponible en version française depuis 1967 (Pichot, Overall et Gorham) est l'instrument de ce type le plus utilisé dans le monde. Revenir sur l'historique de son développement, sur ses caractéristiques et sur les applications qui en ont été faites permet de comprendre les facteurs de ce succès ininterrompu depuis plus de trente ans et, en même temps, de discuter une série de problèmes méthodologiques et pratiques de la psychopathologie quantitative.

On peut faire remonter cette dernière à 1905, date à laquelle un psychologue, Alfred Binet, et un psychiatre, T. Simon, publièrent leur échelle d'intelligence, le premier test mental utilisable en pratique, dont le but était de détecter avec précision le retard mental. Cet instrument de mesure fut le prototype de multiples autres épreuves construites dans des perspectives variées et qui avaient deux caractéristiques fondamentales : elles rassemblent des éléments, ou items, en une échelle; ces éléments sont des stimulus standardisés (dans le test de Binet-Simon des questions-problèmes) auxquels le sujet doit donner une réponse qui est cotée (dans le cas particulier exacte ou non). Le nombre total de réponses cotées positivement dans l'échelle permet de classer le sujet à l'intérieur de la population de référence, et dans l'échelle de Binet-Simon de mesurer son niveau intellectuel.

Parallèlement aux tests mentaux, qui se diversifièrent rapidement en fonction des aspects psychologiques que l'on souhaitait appréhender : intelligence, aptitudes, connaissances, attitudes, intérêts, personnalité, apparut une autre famille d'instruments, ayant la même structure d'échelle, mais dont les items étaient de nature différente. Alors que, dans les tests, le stimulus et le jugement à porter sur la réponse sont rigoureusement standardisés (le rôle de celui qui administre le test et qui cote la réponse est si accessoire qu'aujourd'hui de nombreux tests sont administrés et corrigés informatiquement), dans les échelles appelées d'appréciation, d'observation, ou d'évaluation — en anglais *observer's rating scales*, en abrégé : *rating scales*) — le comportement du sujet est observé dans une situation peu standardisée — par exemple au cours d'un entretien psychiatrique — et ce comportement est quantifié par un observateur suivant des règles préétablies, mais n'ayant nécessairement pas la rigidité de celles qui gouvernent la cotation des réponses aux tests. Du fait de l'intérêt prédominant porté dans la première moitié de ce siècle aux applications scolaires et industrielles, les échelles d'observation alors construites eurent essentiellement ces domaines comme objectifs. Cependant, quelques tentatives isolées de soumettre la

symptomatologie des malades mentaux à la quantification eurent lieu. On peut citer la *Behavior Chart* de Kempf (1914-1915), mais surtout les recherches de T. V. Moore (1930-1933). Cet auteur fut le premier à vouloir, comme il l'écrivait, « déterminer empiriquement la structure de la symptomatologie des psychoses » et, pour ce faire, il construisit une échelle d'observation psychiatrique, l'appliqua à une population de psychotiques, et soumit les résultats à l'analyse factorielle. Ce travail de pionnier souffre évidemment d'insuffisances méthodologiques, mais peut néanmoins être considéré comme le point de départ de tous les efforts ultérieurs.

À partir de 1945 « de nombreux instruments apparurent. Ils furent stimulés par un besoin d'évaluer l'évolution des malades, par l'apparition comme traitement de la lobotomie, et par les progrès faits par les techniques psychométriques, telle l'analyse factorielle » (Lorr *et al.*, 1963). En 1950, J. R. Wittenborn, psychologue à Rutgers University, annonça la mise au point d'un « nouveau procédé pour l'évaluation des malades mentaux hospitalisés ». Il s'agissait d'une échelle dont la version originale fut publiée en 1955 (version française, 1959) qui consistait en 52 items « couvrant un vaste spectre de phénomènes psychopathologiques habituellement rencontrés », chaque item se rapportant à un « symptôme observable », c'est-à-dire soit un comportement du sujet, soit une donnée symptomatique obtenue ou inférée à partir de l'entretien avec celui-ci, dont l'observateur devait apprécier la présence et l'intensité en trois ou quatre degrés. L'analyse factorielle des résultats d'un groupe non sélectionné de malades permit à Wittenborn d'isoler 9 dimensions symptomatiques tels que *anxiété aiguë, hystérie de conversion, syndrome maniaque, syndrome dépressif, excitation schizophrénique*, etc.

À peu près à la même époque, Maurice Lorr, qui dirigeait le laboratoire de recherche psychologique à l'administration américaine des anciens combattants (*Veterans Administration* ou VA), un organisme qui dirigeait un grand nombre de services psychiatriques, construisit une échelle analogue. La *Multidimensional Scale for Rating Psychiatric Patients* (MSRPP) possédait des items analogues à ceux de l'échelle de Wittenborn. Ils pouvaient être cotés après un seul entretien et fournissaient un tableau de l'état du sujet à un moment donné. L'intérêt de Lorr était de nature théorique. Il souhaitait, dans la tradition de Moore, « identifier les syndromes (dimensions) du comportement psychotique ». Pour ce faire, après une révision des travaux existant déjà dans la littérature (Moore, Wittenborn et quelques travaux accessoires), il fit l'hypothèse que 10 dimensions permettaient de couvrir l'ensemble du domaine pathologique. Les items correspondants furent choisis et une forme préliminaire de l'*Inpatient Multidimensional Psychiatric Scale* (IMPS) fut construite et appliquée à un important échantillon de malades. La fidélité fut étudiée, une analyse factorielle pratiquée et, en fonction des résultats, une forme révisée mise au point puis publiée (Lorr *et al.*, 1963). L'échelle est composée de 75 items dont 58 sont cotés en 9 degrés d'intensité ou de fréquence et 17 en oui ou non. Ils sont répartis en 10 sous-échelles factorielles comprenant chacune de 5 à 11 items dont les noms désignent l'élément central des traits correspondants : excitation (EXC), belligérance hostile (HOS), projection paranoïde (PAR), expansivité grandiose (GRN), distorsion perceptive (PCP), intropunitivité anxieuse (INP), ralentissement et apathie (RTD), désorientation (DIS), troubles moteurs (MTR), désorganisation conceptuelle (CNP).

Plusieurs points doivent être soulignés :

1) les items ont été choisis en fonction du fait qu'ils s'adressent à des sujets hospitalisés, pratiquement à l'époque des états psychotiques ou névrotiques graves;

2) ils concernent uniquement des phénomènes observés au cours de l'entretien, aucune information d'autre origine ne devant être utilisée;

3) aucune inférence basée sur une interprétation psychodynamique n'est autorisée. Ainsi on ne doit pas coter le sujet comme «réellement» hostile sous le prétexte qu'il est d'une politesse et d'une déférence exagérées. Ce point, qu'on retrouve dans les échelles ultérieures en particulier dans la BPRS, est d'une particulière importance car, à cette époque, la psychopathologie était aux États-Unis dominée par le psychodynamisme.

Sur le plan métrologique, les caractéristiques de l'échelle (fidélité inter-juges, homogénéité des sous-échelles factorielles) étaient très satisfaisantes. L'échelle fut appliquée à de nombreux malades de la VA et, à partir des résultats, Lorr avec ses collaborateurs put, par l'utilisation de nombreux procédés statistiques de l'analyse multivariée, en particulier de l'analyse factorielle et de l'analyse typologique, apporter une contribution importante à la psychopathologie et à la nosologie quantitative, dans son ouvrage *Explorations in Typing Psychotics* (1966).

Lorsque Wittenborn et Lorr avaient entrepris la construction de leurs échelles, la psychopharmacologie n'était pas encore née. Lorsqu'elle se développa à la fin des années 50, et que bientôt on souhaita quantifier l'effet thérapeutique des nouveaux médicaments, il devint nécessaire de disposer d'instruments spécialement adaptés à cette nécessité, c'est-à-dire fournissant un instantané de la symptomatologie avant le début du traitement et à la fin de l'essai thérapeutique. Certains auteurs utilisèrent tout naturellement les échelles disponibles de Wittenborn et de Lorr. Cette dernière en particulier avait été employée dans les essais thérapeutiques coopératifs multicentriques à large échelle entrepris très rapidement par la VA (ces données ont d'ailleurs en grande partie servi aux travaux typologiques de Lorr). À cette époque, le *National Institute of Mental Health* (NIMH) avait reconnu l'importance de la psychopharmacologie. Il avait créé un centre spécial, le *Psychopharmacology Service Center* devenu ultérieurement la *NIMH Psychopharmacology Research Branch*. Dans ce cadre fut constituée une organisation spéciale, rassemblant des services psychiatriques en un réseau chargé de l'évaluation rapide initiale des nouvelles drogues proposées en nombre croissant par l'industrie, les *Early Clinical Drug Evaluation Units* ou ECDEU. Il apparut rapidement que, pour les ECDEU, l'IMPS avait un défaut pratique : l'emploi de l'instrument demandait beaucoup de temps en raison du nombre élevé d'items et du caractère détaillé des instructions à suivre dans la cotation. Son emploi était certes possible dans des travaux expérimentaux, mais rebutait des praticiens.

C'est dans ces conditions que John E. Overall et D. R. Gorham entreprirent de construire une forme simplifiée de l'IMPS qui fut la BPRS. La simplification porta sur deux points, la présentation des items et leur nombre. Dans l'IMPS, chaque item est décrit avec détails, et contient généralement une liste d'indices (*cues*). La comparaison entre l'item 23 de l'IMPS et l'item 16 de la BPRS, d'orientations identiques, illustre ce point.

IMPS 23 : comparé à une personne normale, à quel degré le malade montre-t-il de l'apathie, de l'indifférence, ou un manque de réponse émotionnelle au cours d'une discussion de ses propres problèmes, sur sa famille, ou sur son milieu? Indices : il ne rit, ne sourit ou ne réagit pas quand on

lui fait une blague; il n'est ni triste, ni irrité; il ne semble pas se soucier de ce qui se passe; il discute de sujets à charge émotionnelle d'une manière plate et détachée.

BPRS 3 : réduction du tonus émotionnel, impression d'un manque de sensibilité et de participation affective.

En outre, l'échelle de 9 adverbes d'intensité ou de fréquence qui sert à la cotation de la majorité des items de l'IMPS fut restreinte dans la BPRS à 7. Le second aspect significatif de la simplification consista à ramener le nombre d'items de 75 à 16. Peut-être le point le plus important fut-il la qualité de ce choix. Overall utilisa les analyses factorielles de l'IMPS pour sélectionner des items dont les contenus appartenaient aux différentes dimensions isolées de manière à couvrir autant que possible l'ensemble du champ à explorer : la symptomatologie des malades hospitalisés traitables par pharmacothérapie. Dans ces conditions, la BPRS se révéla un instrument d'utilisation simple et rapide, le psychiatre établissant sa cotation après un examen standardisé d'une durée de 15 à 20 minutes. Elle fut acceptée avec enthousiasme par le NIMH, son usage étant recommandé par les directeurs successifs des programmes psychopharmacologiques, Jonathan O. Cole, Jerome Levine et Nina R. Schooler. Elle fut incorporée dans les manuels d'épreuves standard des ECDEU (McGlashan, 1966; Guy, 1976) et utilisée très largement dans le monde entier. Elle fut introduite en France en 1967 (Pichot, Overall et Gorham).

Depuis qu'elle a été publiée, la BPRS a fait l'objet de nombreuses propositions de modification, portant sur le nombre et la formulation des items.

La forme originale de 1962 contenait 16 items, destinés à la mesure du changement et, du fait que les essais concernés se rapportaient surtout à l'effet des neuroleptiques chez les schizophrènes, ceci avait influencé leur choix, d'autant que la *Rating Scale for Depression* publiée en 1960 par Max Hamilton allait bientôt être l'instrument privilégié pour les essais d'antidépresseurs. Vers 1965, Overall estima qu'il serait intéressant d'utiliser également l'instrument, malgré sa brièveté, dans la tradition des travaux de Lorr, pour des recherches sur la classification des troubles mentaux. Il créa dans ce but deux nouveaux items *excitation* et *désorientation* qui couvraient deux des 10 facteurs de l'IMPS (EXC et DIS), non représentés dans la BPRS initiale. La version à 18 items est celle qui est recommandée par les ECDEU et qui est généralement utilisée dans le monde. Toutefois, comme l'a fait remarquer avec humour Overall (1988), la bibliographie de la plupart des articles utilisant l'échelle à 18 items continue à donner comme référence le travail de 1962 où est décrite l'échelle à 16 items.

Depuis 1965, Overall a été très réticent pour modifier le nombre d'items. Il écrit en 1988 : «La BPRS n'est certes pas parfaite et l'adjonction de certains nouveaux items pourrait augmenter son utilité comme instrument de description de l'ensemble de la psychopathologie observable ... si l'on devait ajouter des items à la liste, "exaltation de l'humeur" et "affect inapproprié" seraient les premiers pris en considération». Mais il ajoute qu'il s'est refusé d'un point de vue pragmatique, à suivre cette direction, car il se déclare «fermement convaincu que la BPRS n'aurait pas survécu si on avait continué à la modifier».

Toutefois, Overall a participé en 1971 à une tentative de préparer, à partir de la BPRS, une échelle beaucoup plus longue (42 items) constituée des 18 items de la BPRS classique auxquels étaient ajoutés l'item «humeur expansive», 12 items empruntés à la «*Minor Symptom Scale*» mise au point par Overall et Tupin en 1966 et 11 items construits par Pokorny et Overall en 1966 et 1972. Cette échelle explore ainsi, outre les domaines psychotiques, des symptômes appartenant aux domaines

des déficits cognitifs, des troubles de la personnalité, des manifestations névrotiques et des comportements toxicomaniaques. Bien que cette échelle se soit, dans sa version française, montré intéressante du fait même de l'élargissement du champ exploré (Pichot, Samuel-Lajeunesse, Lebeaux, 1973), elle n'a pas eu d'extension internationale.

Parmi les versions à plus de 18 items préparés indépendamment de Overall, il faut citer la BPRS (E) (Lukoff *et al.*, 1986) destinée spécifiquement à l'étude de la réhabilitation des schizophrènes. Elle contient six items supplémentaires concernant la bizarrerie du comportement, la négligence de soi, la suicidalité, l'humeur expansive, l'hyperactivité motrice et la distractibilité.

À l'inverse des tentatives précédentes, il existe un essai de réduction dû à Per Bech (1988). Overall avait (1976) indiqué l'utilité de la « note totale de pathologie » obtenue en additionnant les notes aux 18 items. Bech a souhaité obtenir une telle note totale, mais se référant uniquement aux symptômes des malades schizophrènes, en d'autres termes de constituer une sous-échelle mesurant la dimension « intensité de la symptomatologie schizophrénique ». Il a employé deux méthodes pour sa construction. La première est l'analyse de jugement de Hammond qui consiste à demander à un groupe de psychiatres experts de coter de 0 à 10 l'importance (le poids) de chaque item dans la perspective envisagée. La seconde est la méthode probabiliste de Rasch, qui repose sur un modèle développé initialement pour la construction de tests d'intelligence et de connaissances (Rasch, 1960). Schématiquement, ce modèle permet, par une analyse statistique complexe, de déterminer dans une échelle quels items, classés dans un certain ordre, mesurent une dimension latente de manière à ce que, pour cette population d'items et pour une population de personnes, la position de chaque item soit invariante quel que soit l'échantillon de sujets et que la position de chaque sujet soit invariante dans différents échantillons d'items. Les résultats d'une telle analyse logistique ont confirmé à Bech que les 10 items de la BPRS qu'il avait isolés par l'analyse du jugement comme mesurant l'intensité de la symptomatologie schizophrénique permettaient une mesure d'une dimension latente, et que, par conséquent, la hiérarchie des 10 items se maintenait quel que soit le groupe de schizophrènes considéré.

Il existe également des tentatives de modifier la formulation des items et les instructions sur les conditions d'observation et de cotation dont le but est généralement d'améliorer la fidélité inter-juges. Dans le manuel des ECDEU pour la version à 18 items, Overall a donné les commentaires suivants sur son mode d'emploi :

1) « Les évaluations sont basées sur les informations obtenues au cours d'un entretien clinique d'environ 20 minutes ».

2) « Il est recommandé que chaque malade soit l'objet d'un entretien et évalué indépendamment par deux observateurs professionnels » mais il ajoute en 1976 que « l'avantage obtenu par l'existence de deux évaluations indépendantes n'est pas aujourd'hui considéré comme aussi important qu'il l'était autrefois ».

Le manuel des ECDEU contient en outre des commentaires détaillés, pouvant atteindre 15 lignes pour certains items, sur les éléments à prendre en considération pour la cotation. D'autre part, le manuel a ajouté, pour des raisons d'homogénéisation de la batterie d'épreuves, aux 7 degrés d'intensité ou de fréquence initiaux la possibilité de coter un item « non évalué ». À part ces modifications approuvées par Overall, les autres peuvent être classées en quatre catégories :

1) *Diminution du nombre de degrés de cotation.* Bech, dans la version qu'il a publiée de la BPRS dans son « Mini-Compendium », n'utilise que 5 degrés. Bigelow les a réduits à 4, mais d'une manière complexe. Les 7 degrés originaux sont conservés mais par exemple 2 (originalement : très léger) et 3 (originalement : léger) sont définis conjointement par léger, le choix étant laissé au cotateur de choisir entre 2 et 3.

2) *Modification de la définition des degrés.* La technique de la BPRS consiste en l'emploi d'adverbes. Elle repose sur de nombreux travaux de psychologie expérimentale qui ont montré que les adverbes possédaient dans le langage une valeur de « multiplicateur » qui est stable quels que soient le sujet qui l'emploie et le phénomène qu'il qualifie. Ceci donne la possibilité de sélectionner des séries graduées d'adverbes, chacun étant séparé du précédent et suivant par une distance psychologique égale. Une autre technique consiste à définir chaque degré par une phrase ou un paragraphe précisant les comportements qui permettent de considérer ce degré comme approprié. Ces phrases sont généralement dénommées « descripteurs » ou « points d'ancrage ». Cette substitution de phrases aux adverbes a été proposée par Woerner *et al.* (1988), lorsque l'échelle a été utilisée par le NIMH dans son étude collaborative sur les stratégies de traitement de la schizophrénie. La forme ainsi modifiée a été dénommée BPRS(A) (A pour ancrée) et est en fait très différente de la BPRS initiale car elle inclut en outre une modification de la formulation des items. D'autre part, les auteurs ont préparé un *Guide d'entretien.* Raskin (1988) a fait remarquer que, en définissant chaque « point d'ancrage » par des comportements extrêmement limités (par exemple l'intensité « modérée » de l'item « repli émotionnel » est notée si le sujet « a peu de contacts avec les yeux, mais pourtant paraît impliqué dans l'entretien et répond de manière appropriée à toutes les questions »), on court le risque qu'il soit différent de celui défini par un adverbe.

3) *Définition des conditions d'observation.* Contrairement aux consignes initiales d'Overall, Raskin (1969) demande à l'observateur « de décrire l'état du malade comparé à celui d'une personne normale des mêmes âge et sexe ». Woerner *et al.* indiquent pour la BPRS(A) que la cotation est basée sur « les informations concernant la dernière semaine » alors que pour Bigelow, dans sa révision, elle l'est sur « le comportement du malade uniquement le jour de l'examen ». On peut penser que l'idée initiale d'Overall était de s'aligner sur les consignes données par Lorr pour l'IMPS.

4) *Introduction d'entretiens plus ou moins standardisés.* Dans la pensée d'Overall, l'entretien psychiatrique devait être flexible. Il a indiqué qu'un avantage de la BPRS était que « les personnes conduisant l'entretien sont rapidement capables d'avoir à l'esprit les différents contenus d'items sans avoir à suivre une succession déterminée de questions ». La publication d'entretiens psychiatriques plus ou moins structurés s'est multipliée en psychiatrie en liaison avec l'apparition de nouvelles nosologies, comme le DSM-III et la CIM-10. C'est dans cette atmosphère que l'on a tenté de mettre au point de tels instruments pour guider les utilisateurs de la BPRS. Ainsi, Tarell et Schulz, dans le cadre du NIMH, ont préparé un « interview structuré » consistant en une série de questions précises à poser au malade. La valeur de cet instrument est difficile à juger, car il concerne une forme modifiée de la BPRS et qui de plus est destinée à être utilisée par les infirmières. Overall, malgré sa réticence à sacrifier la flexibilité, a fait un pas dans cette direction. Il a, en 1986, discuté avec Hollister de la technique de conduite de l'entretien, des éléments sur lesquels devrait être basée l'évaluation de chaque item, et donné une liste de questions de nature

générale à poser éventuellement au malade. En 1988, il est revenu sur le sujet et a donné une série de 12 questions, mais en insistant sur le fait que plusieurs de ces questions se rapportent au même item et qu'inversement chaque question peut donner des indications sur plusieurs items.

Le problème de telles modifications, dont certaines ont certainement des avantages (extension de l'aire pathologique couverte et surtout amélioration de la fidélité inter-juges), a été bien résumé par Raskin : « Le chercheur doit décider soit d'utiliser la BPRS avec des propriétés psychométriques qui peuvent paraître limitées suivant les standards actuels mais qui possèdent des normes étendues et une aptitude bien démontrée à détecter les effets des traitements ou si l'on préfère utiliser de nouveaux instruments possédant par exemple un interview structuré et/ou des points d'ancrage définis. »

Ces propriétés psychométriques évoquées par Raskin correspondent aux quatre grandes variétés de qualités métrologiques qui caractérisent tout instrument : la sensibilité, l'homogénéité, la fidélité et la validité.

La sensibilité concerne la finesse de la discrimination permise par l'instrument entre degrés de la propriété mesurée. Elle est directement liée au nombre d'items constituant l'échelle et au nombre de degrés de cotation de chaque item. Dans les instruments, on est amené à un compromis : une échelle contenant beaucoup d'items sera sensible, mais en pratique difficile à utiliser, de la même manière que si l'on prétendait graduer un décamètre en millimètres. La BPRS a, par construction, une sensibilité inférieure à l'IMPS. À son succès a contribué le fait que cette sensibilité était suffisante pour son application principale, la mesure du changement dans les essais thérapeutiques, comme le démontrait l'existence de différences statistiquement significatives entre notes pré- et posttraitement. Un auteur comme Bech a estimé que la sensibilité de la BPRS était facticement surestimée : du fait de l'existence de 7 degrés de notation, la note totale peut varier de 18 à 126, soit une marge de 108 points. Il a proposé de réduire le nombre de degrés à 4 en estimant qu'un observateur était incapable de choisir de manière fidèle entre un nombre supérieur. Il faut faire remarquer à ce sujet qu'une telle démarche ne peut que réduire la sensibilité. Le seul risque que l'on court en utilisant 7 degrés est de l'augmenter.

L'homogénéité est la qualité qui fait que les différents items mesurent tous la même qualité. Si elle est fondamentale dans une épreuve qui prétend mesurer une dimension (l'intelligence, l'intensité de la dépression), son étude n'a pas de sens dans une épreuve comme la BPRS car, dans sa construction, aucune hypothèse n'a été faite sur l'existence d'une dimension unique sous-jacente à l'ensemble des manifestations psychopathologiques observées. Par contre, il est possible, comme l'a fait Bech en utilisant le modèle de Rasch, de démontrer que certains items de la BPRS constituent une échelle homogène d'intensité de la symptomatologie schizophrénique. De la même manière, les sous-échelles isolées dans la BPRS par l'analyse factorielle ont, par construction, une homogénéité élevée.

La fidélité (*reliability*) se définit par la stabilité du résultat au cours de mesures répétées. Dans le cas d'une échelle d'évaluation, cette répétition peut être le fait de deux (ou plus) observateurs effectuant indépendamment, mais au même moment et dans les mêmes conditions, leur appréciation (fidélité inter-juges, *inter-rater reliability*) ou correspondre à des mesures plus ou moins éloignées dans le temps (stabilité temporelle). Il existe différentes méthodes de calcul de la stabilité temporelle mais ceci est sans objet pour la BPRS qui vise à donner une évaluation à un

moment donné d'un état qui, par hypothèse, n'est pas stable : l'état pathologique qui évolue spontanément ou sous l'influence d'un traitement. La seule fidélité importante dans ce cas, comme dans celui de toutes les échelles d'observation de la symptomatologie, est la fidélité inter-juges. De nombreuses recherches ont été faites sur le sujet, mais l'interprétation des résultats, exprimés par le calcul de corrélations entre notes attribuées par deux observateurs aux mêmes sujets (le coefficient exprime la proportion de variance non-erreur de la mesure), est difficile. Le coefficient varie, pour une même valeur absolue de la fidélité inter-juges, en fonction de l'étalement des notes dans le groupe étudié et du nombre d'items considéré. Dans une revue générale de 1980, Hedlund et Vieweg ont conclu que « lorsque la BPRS est correctement utilisée, la fidélité inter-juges est généralement satisfaisante ». Toutefois, Hafkenscheid a noté que certaines publications récentes étaient moins optimistes et il attribue le fait à ce que les coefficients de corrélation élevés avaient été obtenus « dans des conditions optimales : observateurs bien entraînés aux conditions de recherche, ayant une grande pratique de l'instrument pendant une période prolongée, ayant probablement développé un ensemble largement implicite de définitions et de points d'ancrage. » Les différentes tentatives de modifier la BPRS ont eu pour but principal d'améliorer cette qualité métrologique.

La *validité* est la qualité métrologique sans doute la plus importante d'une épreuve. Celle-ci est valide si elle mesure ce qu'elle prétend mesurer. Comme l'indique Cronbach, les instruments psychométriques ont pour but de permettre des décisions et il est possible qu'une épreuve soit valide dans une certaine perspective et non dans une autre. On distingue habituellement quatre types de validité :

– la *validité prédictive*, qui répond à la question : les résultats à l'épreuve prédisent-ils un comportement important ultérieur du sujet?

– la *validité concourante* (*concurrent validity*) qui répond à la question : les résultats à l'épreuve permettent-ils une estimation d'un certain comportement actuel du sujet?

– la *validité de contenu* qui répond à la question : comment les résultats à l'épreuve peuvent-ils être expliqués psychologiquement?

Il existe des méthodes appropriées à la mesure de chaque variété de validité. Dans le cas de la BPRS qui a été construite pour mesurer le changement de la symptomatologie sous l'influence du traitement, un indice de la validité prédictive est la diminution des notes à l'application postthérapeutique.

La validité concourante est souvent mesurée en administrant aux mêmes sujets l'épreuve à valider et une épreuve déjà utilisée dont on considère qu'elle mesure correctement le même phénomène. Cette pratique très répandue est de valeur douteuse car, si les résultats des deux épreuves sont faiblement corrélés, il est impossible de savoir si ceci est dû à une supériorité ou à une infériorité de la nouvelle épreuve par rapport à l'ancienne. Dans le cas des instruments telle que la BPRS, la validité concourante se mesure communément en corrélant le résultat à un jugement clinique. Ainsi, la note totale reflétant la gravité de la symptomatologie, la forte corrélation constatée entre cette note et un jugement global de cliniciens sur cette gravité sont des indices de validité.

La validité de contenu est basée sur la comparaison logique entre l'énoncé des items et le comportement devant être mesuré. Par définition, les échelles telle que la BPRS ont une forte validité de contenu dans la mesure où la rédaction des items repose sur la description que les cliniciens font des éléments psychopathologiques correspondants.

– La *validité de construct* se mesure en formulant des hypothèses et en les testant par des procédés appropriés. Dans le cas particulier, un des procédés les plus employés est l'analyse factorielle qui permet de tester l'hypothèse selon laquelle la symptomatologie mesurée peut être structurée de manière interprétable. L'analyse factorielle est une technique d'analyse statistique multivariée qui vise à représenter de manière plus économique avec une perte minimale d'information (de variance) la matrice d'intercorrélations entre items appliqués à une population d'individus : elle substitue aux N items un nombre n plus petit de dimensions, chacune étant définie par une combinaison appropriée d'items co-variants. Les facteurs sont ainsi les dimensions sous-jacentes s'exprimant dans les items. Sur le plan technique, il existe de nombreux procédés qui donnent tous des solutions mathématiquement correctes. Le choix entre eux est déterminé par trois éléments. L'un est pratique, la disponibilité facile de programmes de calculs. Les deux autres sont théoriques. On tend à choisir la solution qui est :

a) la plus simple, suivant le principe scholastique de parcimonie du rasoir d'Occam : «*Entia non sunt multiplicata praeter necessitatem*» ;

b) la plus compréhensible psychologiquement.

La méthode aujourd'hui la plus employée comporte deux phases. La première est l'analyse en composantes principales. Dans celle-ci, on détermine la dimension qui explique le maximum de variance de la matrice, puis celle, orthogonale à la première, qui explique le maximum de la variance restante, et ainsi de suite. En théorie, on peut extraire autant de composantes qu'il y a d'items. Mais dans la matrice, chaque coefficient de corrélation contient une part de variance due aux erreurs de mesure, et à partir d'un certain point, les composantes principales extraites ne correspondent plus qu'à cette variance erreur. On doit décider du nombre de composantes suffisant pour expliquer la variance «vraie», nombre nécessairement inférieur au nombre d'items. Les composantes principales retenues apportent ainsi une simplification de la matrice originale (dans la nouvelle matrice, la corrélation entre item et composante est appelée saturation). Les composantes principales ne sont généralement pas interprétables psychologiquement.

On procède alors au second temps de l'analyse, la rotation des composantes (le terme correspond au fait que les n composantes peuvent être représentées géométriquement dans un hyper-espace à n dimensions) de manière à ce que la solution soit aussi simple que possible, c'est-à-dire que le nombre de saturations nulles des items soit maximal. Les composantes après rotation sont appelées *facteurs*. La méthode aujourd'hui la plus employée est la rotation dite *Varimax de Kaiser*. Elle maintient les facteurs orthogonaux.

De nombreuses autres méthodes existent, toutes donnant des solutions mathématiquement correctes. Ainsi Lorr a, dans ses travaux, utilisé une méthode dans laquelle les axes deviennent obliques les uns par rapport aux autres, ce qui fait que les notes factorielles sont intercorrélées, estimant que la structure était plus simple. Pratiquement tous les travaux sur la BPRS ont utilisé la méthode Varimax bien que Overall et Klett (1972) aient présenté une solution oblique. On ne peut s'attendre à ce que toutes les analyses factorielles donnent des résultats identiques car les matrices d'intercorrélation de départ dépendent en partie de la population sur laquelle elles ont été calculées.

L'analyse à laquelle il est le plus souvent fait référence est celle publiée par l'ECDEU, basée sur les résultats de 3596 schizophrènes. Elle décrit 5 facteurs après

rotation Varimax, dénommés d'après la signification des items ayant une forte saturation dans chacun d'eux : *anxiété-dépression* (ANDP) (4 items), *troubles de la pensée* (THOT) (4 items), *activation* (ACTV) (3 items), *méfiance hostile* (HOST) (3 items). Dans cette analyse, chaque item de la BPRS a une saturation significative dans un facteur et un seul. Une autre solution, basée sur la comparaison de multiples analyses, a été proposée (Overall, Hollister, Pichot, 1967). Elle retient quatre facteurs fondamentaux : *troubles de la pensée*, *méfiance hostile délirante*, *repli-ralentissement psychomoteur*, *dépression anxieuse*. Overall estime que la structure est en fait composée de quatre dimensions bipolaires sous-jacentes à la symptomatologie des troubles mentaux majeurs. Les items de la BPRS mesurent un pôle du continuum de chaque dimension. Dans le cas de la désorganisation conceptuelle, l'autre pôle n'appartient pas à la pathologie, mais les trois autres peuvent être formulés : intropunitivité contre extrapunitivité, repli-ralentissement contre agitation-excitation, dépression contre élation. Les solutions à cinq et à quatre facteurs ont des rapports étroits, et le contenu de chacun des facteurs est susceptible d'interprétation psychologique facile, ce qui constitue un argument en faveur de la validité de construct de la BPRS.

Le développement d'une épreuve ne s'arrête pas à la détermination de ses qualités métrologiques. Il requiert l'accumulation d'un corpus de données à partir de l'analyse des résultats. Il est peu d'échelles pour lesquelles ce corpus soit aussi important que la BPRS (Overall, 1974, 1976; Overall et Hollister, 1986; Overall et Klett, 1972; Hedlund et Vieweg, 1980). On ne peut que rappeler trois grands chapitres.

Établissement de profils moyens caractéristiques de différentes catégories de différents troubles mentaux

Deux approches ont été employées. La première a consisté à utiliser des malades réels. La seconde (méthode des stéréotypes) demande à une série de psychiatres de coter la BPRS comme s'ils se trouvaient en face d'une malade présentant la symptomatologie qu'ils considèrent comme typique d'un diagnostic psychiatrique particulier. Ces travaux ont été entrepris avant 1980, à une époque où n'existait pas de nosologie internationalement utilisée (type DSM-III ou CIM-10) dans laquelle chaque catégorie est définie par des critères diagnostiques précis. La mise au point de stéréotypes dans différents pays (États-Unis, France, Allemagne, etc.) a permis de démontrer que les mêmes termes (par exemple schizophrénie paranoïde) recouvraient des concepts nosologiques ayant des symptomatologies différentes. Une technique statistique de calcul des distances entre stéréotypes moyens a précisé la nature de ces différences qui constituaient un handicap considérable dans la communication internationale, par exemple concernant les indications des drogues psychotropes.

Détermination statistique de types syndromiques

Inaugurée par Lorr avec l'IMPS, cette recherche vise à déterminer si, dans une population hétérogène de malades, on peut définir des sous-groupes homogènes indépendants de sujets caractérisés par une symptomatologie identique ou très voisine — ce qui correspond à la définition du terme syndrome en médecine. (Du fait que les psychologues utilisent le mot syndrome pour les dimensions isolées par l'analyse factorielle et pour ces sous-groupes homogènes de sujets, il est nécessaire de préciser s'il s'agit de « syndromes-traits » ou de « syndromes-types ».) La méthode appropriée

à la découverte de ces sous-groupes est l'analyse typologique dont il existe plusieurs variétés (analyse en clusters, analyse factorielle inversée). Overall a isolé trois sous-groupes dépressifs : *dépression anxieuse*, *dépression hostile*, *dépression repliée-ralentie*, et trois sous-groupes dans les psychoses délirantes : *syndrome délirant hostile-méfiant*, *syndrome avec désorganisation de la pensée et repli*, *syndrome avec troubles florides de la pensée*. Il a démontré la validité de ces syndromes, dont les profils moyens à la BPRS ont été publiés, par leur sensibilité différentielle à la thérapeutique médicamenteuse.

Diagnostic par ordinateur

Il est possible de calculer la distance entre le profil d'un sujet et les profils moyens de différentes catégories de malades. Ces profils moyens peuvent avoir été obtenus comme indiqué ci-dessus soit à partir de malades réels, soit à partir des représentations stéréotypes de psychiatres, soit par analyse typologique. De toute manière, on détermine le profil moyen dont le profil du sujet est le plus proche, et qui constitue donc son diagnostic le plus probable. Bien entendu, il s'agit là d'un diagnostic basé uniquement sur les caractéristiques symptomatiques actuelles telles qu'elles sont cotées par la BPRS.

Certaines de ces applications ont cherché à résoudre des problèmes théoriques : correspondance entre nosologies nationales, établissement d'une nosologie syndromique originale empirique par l'emploi des statistiques multivariées, mais elles ont conservé à l'arrière-plan des préoccupations relatives aux applications psycho-pharmacologiques.

Des échelles autres que la BPRS lui sont aujourd'hui préférées dans certains domaines des essais thérapeutiques comme dans celui de la dépression ou dans ceux des états névrotiques, car sa sensibilité a été jugée insuffisante, dans le premier cas, du fait du faible nombre d'items concernés et parce que, dans le second, elle n'avait pas, d'ailleurs délibérément, inclus les items appropriés. Malgré ces réserves, le contraste entre la brièveté de l'échelle et la qualité des résultats obtenus dans la pratique aussi bien que dans la recherche théorique à laquelle elle a servi de point de départ est frappant. Ceci doit être rapporté à sa valeur technique, résultat d'une technique de développement correcte, tenant compte de manière optimale des exigences méthodologiques et des nécessités pratiques.

BIBLIOGRAPHIE*

* Les références précédées d'un astérisque correspondent soit aux principales références historiques, soit à celles qui contiennent une bibliographie importante.

ANDERSEN J., LARSEN J. K., SCHULTZ V., *et al.* — The Brief Psychiatric Rating Scale Dimensions of Schizophrenia. Reliability and Construct Validity. *Psychopathology*, *22*, 168-176, 1989.

BECH P., KASTRUP M., RAFAELSEN O. J. — *Mini-compendium of Rating Scales for States of Anxiety, Depression, Mania and Schizophrenia with Corresponding DSM-III Syndromes.* Acta Psychiatr. Scand. 73 (Suppl. 329) Copenhagen, 1986. Trad. franç., Masson, Paris, 1989.

BECH P., LARSEN J. K., ANDERSEN J. — The BPRS : Psychometric Developments. *Psychopharmacol. Bull.*, *24*, 118-121, 1988.

DINGEMANS P. M., FROHN-DE WINYER M. L., BLEEKER J. A. C., RATHOD P. — A Cross-Cultural Study of the Reliability and Factorial Dimensions of the Brief Psychiatric Rating Scale (BPRS). *Psychopharmacol.*, *80*, 190-191, 1983.

EISELE E., GEX-FABRY M., BALANT-GORGIA A. E., BALANT L., GARRONE G. — Rationale for BPRS Use in Routine Clinical Practice : Quantitative Assessment for Psychopathology, Consistent with Clinical Sense. *Eur. Psychiatry*, *6*, 261-268, 1991.

FLEMENBAUM A., ZIMMERMANN R. I. — Inter- and Intra-rater Reliability of the Brief Psychiatric Rating Scale. *Psychol. Rep.*, *36*, 783-792, 1973.

GABBARD G. O., KENNEDY L. L., DEERING C. D., *et al.* — Inter-rater Reliability in the Use of the Brief Psychiatric Rating Scale. *Bull. Menninger Clin.*, *51*, 519-531, 1987.

GINOLHAC T., PICHOT P. — La fidélité inter-juges de la BPRS 42. *Ann. Médico-Psychol.*, *113 (1)*, 147-151, 1975.

GUY W. — *ECDEU Assessment Manual for Psychopharmacology. Revised 1976.* Rockville : NIMH Psychopharmacology Research Branch, 1976.

HAFKENSCHAID A. — Psychometric Evaluation of a Standardized and Expanded Brief Psychiatric Rating Scale. *Acta Psychiatr. Scand.*, *84*, 294-300, 1991.

* HEDLUND J. L., VIEWEG B. M. — The Brief Psychiatric Rating Scale. A Comprehensive Review. *J. Operational Psychiatry*, *11*, 45-65, 1980.

KEMPF E. J. — The Behavior Chart in Mental Disease. *Amer. J. Insanity*, *71*, 761-772, 1914-1915.

LORR M., KLETT C. J., MCNAIR D. M., LASKY J. J. — *Inpatient Multidimensional Psychiatric Scale. Manual.* Palo Alto : Consulting Psychology Press, 1963.

* LORR M. (Ed.) — *Explorations in Typing Psychotics.* Pergamon Press, Oxford, 1966.

* LORR M., KLETT C. J., MCNAIR D. M. — *Syndromes of Psychosis.* Pergamon Press, Oxford, 1963.

MCGLASHAN T. (Ed) — *The Documentation of Clinical Drug Trials.* Rockville National Institute of Mental Health.

MOORE T. V. — The Empirical Determination of Certain Syndromes Underlying Praecox and Manic-Depressive Psychoses. *Amer. J. Psychiatry*, *86*, 719-738, 1930.

* MOORE T. V. — The Essential Psychoses and their Fundamental Syndromes. *Stud. Psychol. & Psychiatr.*, *3*, 1-28, 1933.

* OVERALL J. E. — *The Brief Psychiatric Rating Scale in Psychopharmacology Research.* Psychometric Laboratory Reports n° 29. Galveston, University of Texas, 1972.

* OVERALL J. E. — The Brief Psychiatric Rating Scale in Psychopharmacology. In PICHOT P. (Ed). *Psychological measurements in psychopharmacology.* Modern Problems of Pharmacopsychiatry, 7. Karger, Basel, 1974.

* OVERALL J. E. — The Brief Psychiatric Rating Scale in Psychopharmacological Research. In GUY W. (Ed.) *Op. cit,* 1976.

OVERALL J. E. — The Brief Psychiatric Rating Scale (BPRS) : Recent Developments in Ascertainment and Scaling. *Psychopharmacol. Bull.*, *24*, 97-99, 1988.

* OVERALL J. E., GORHAM D. R. — The Brief Psychiatric Rating Scale. *Psychol. Rep.*, *10*, 799-812, 1962.

OVERALL J. E., GORHAM D. R. — A Pattern Probability Model for Classification of Psychiatric Patients. *Behav. Sci.*, *8*, 108-116, 1963.

OVERALL J. E., HOLLISTER L. E. — Computer Procedures for Classification of Psychiatric Patients. *J. Am. Med. Ass.*, *187*, 583-588, 1963.

OVERALL J. E., HOLLISTER L. E. — Studies in Quantitative Approach to Psychiatric Classification. In KATZ M. M., COLE J.E., BARTON W. (eds). *The Role of Methodology of Classification in Psychiatry and Psychopathology.* Chevy Chase : U.S. Department of Health, Education and Welfare, 1965.

* OVERALL J. E., HOLLISTER L. E. — Assessment of Depression using the Brief Psychiatric Rating Scale. In SARTORIUS N., BAN T. (Eds). *Assessment of Depression.* Springer, Heidelberg, 1986.

OVERALL J. E., HOLLISTER L. E., PICHOT P. — Major Psychiatric Disorders. A Four Dimensional Model. *Arch. Gen. Psychiatry, 16,* 146-151, 1967.

* OVERALL J. E., KLETT C. J. — *Applied Multivariate Analysis.* McGraw Hill, New York, 1972.

PICHOT P., BAILLY R., DESMEDT D. — Les méthodes de contrôle des thérapeutiques. II. Vérification de la validité inter-correcteur de l'échelle de Wittenborn. *Rev. Psychol. Appl., 9,* 117-118, 1959.

PICHOT P., BAILLY R., OVERALL J. E. — Les stéréotypes diagnostiques des psychoses chez les psychiatres français. Comparaison avec les stéréotypes américains. In *Proceedings of the Vth International Congress of the CINP.* Excerpta Medica Foundation Congress Series n° 129. Excerpta Medica Foundation, Amsterdam, 1967.

PICHOT P., DREYFUS J. F., NICOLAS M. — Étude de l'accord entre psychiatres sur la symptomatologie enregistrée par échelles d'appréciation de malades présentés par magnétoscope. *Ann. Médico-Psychol., 133 (1),* 152-159, 1975.

PICHOT P., LEMPERIERE T., BAILLY F. — Les méthodes de contrôle des thérapeutiques. I - Étude sur la fidélité de l'échelle de Wittenborn. *Rev. Psychol. Appl., 8,* 43-36, 1958.

* PICHOT P., OVERALL J. E., GORHAM D.R. — *Échelle abrégée d'appréciation psychiatrique (Brief Psychiatric Rating Scale) - BPRS.* Éditions du Centre de psychologie appliquée, Paris, 1967.

* PICHOT P., OVERALL J. E., SAMUEL-LAJEUNESSE B., DREYFUS J. F. — Structure factorielle de l'échelle abrégée d'appréciation psychiatrique (BPRS). *Rev. Psychol. Appl., 19,* 217-232, 1969.

* PICHOT P., SAMUEL-LAJEUNESSE B., LEBEAUX A. M. — Étude d'une nouvelle forme expérimentale de la BPRS. *Ann. Médico-Psychol., 131 (2),* 254-263, 1973.

PULL C. B., PICHOT P., OVERALL J. E. — A Statistical Model of Psychiatric Practice in France. *Comprehensive Psychiatry, 20,* 27-39, 1979a.

PULL C. B., PICHOT P., OVERALL J. E. — Établissement à l'aide d'un programme d'ordinateur de la correspondance française de certains diagnostics américains. *Ann. Médico-Psychol., 137,* 59-70, 1979b.

PULL C. B., PICHOT P., OVERALL J. E., PULL M. C. — Une classification syndromique des maladies mentales. II. Homogénéité de 13 stéréotypes syndromiques. *Ann. Médico-Psychol., 135 (1),* 102-107, 1977a.

PULL C. B., PICHOT P., OVERALL J. E., PULL M. C. — Une classification syndromique des maladies mentales. III. Comparaison des stéréotypes français avec les stéréotypes américains. *Ann. Médico-Psychol., 135 (1),* 109-117, 1977b.

RASCH G. — *Probabilistic Models for Some Intelligence and Attainment Tests.* The Danish Institute for Educational Research, Copenhagen, 1960.

RASKIN A. — Discussion : Recent Developments in Ascertainment and Scaling of the BPRS. *Psychopharmacol. Bull., 24,* 122-124, 1988.

RHOADES J. D., SCHULZ S.C. — The Semi-Structured BPRS Interview and Rating Guide. *Psychopharmacol. Bull., 24,* 101-104, 1988.

TARELL J. D., SCHULZ S.C. — Nursing Assessment Using the BPRD : A Structured Interview. *Psychopharmacol. Bull., 24,* 105-111, 1988.

* WITTENBORN J. R. — A New Procedure for Evaluating Mental Hospital Patients. *J. Consult. Psychol., 14,* 500-501, 1950.

* WITTENBORN J. R. — *Wittenborn Psychiatric Rating Scales Manual.* Psychological Corporation, New York, 1955.

WITTENBORN J. R. — Wittenborn Rating Scales. In GUY W. (Ed.) *Op. cit.*, 1976.

* WOERNER M. G., MANNUZZA S., KANE J. M. — Anchoring the BPRS : An Aid to Improved Reliability. *Psychopharmacol. Bull.*, *24*, 112-117, 1988.

3

APPROCHE POLYDIAGNOSTIQUE : INTÉRÊTS ET MÉTHODES DANS LA SCHIZOPHRÉNIE

par S. DOLLFUS

INTÉRÊTS

Alors que la première description d'un « dément précoce » a été faite par B. Morel en 1860, les systèmes de critères diagnostiques pour la schizophrénie ne se sont réellement développés que depuis ces deux dernières décennies bien après les observations princeps de Kraepelin, de Bleuler puis de celles de Schneider.

La détermination à établir des définitions opérationnelles dans les années 70 s'est concrétisée par la parution en 1972 de critères opérationnels pour 14 catégories diagnostiques (Feighner *et al.*, 1972) dont la schizophrénie. D'autres se développeront par la suite; les principaux étant aux États-Unis, les critères diagnostiques pour la recherche (RDC – Spitzer *et al.*, 1978), le DSM-III (1980) et le DSM-III-R (1987), en France, les critères de Pull *et al.* (1987-a, b) et dans les pays germaniques, les critères de recherche de Vienne (Berner *et al.*, 1987). De plus, la classification internationale des maladies mentales (CIM-10, 1992) a récemment été mise à jour.

Ces systèmes diagnostiques comprennent des critères monothétiques ou polythétiques :

– Par critères monothétiques, on entend des critères obligatoires. Ils peuvent être uniques ou multiples. La présence de ces seuls critères dans un système diagnostique aboutit à la sélection de populations homogènes.

– Par critères polythétiques, on entend des critères facultatifs. La plupart des critères inclus dans les classifications diagnostiques sont polythétiques, c'est-à-dire qu'ils renferment au choix au moins un des deux, trois (ou plus) items. C'est le cas par exemple des items inclus dans le critère Al de la schizophrénie selon le DSM-III-R pour lesquels deux sur les 5 items sont requis.

La plupart des systèmes diagnostiques renferment les deux types de critères. Par exemple pour le diagnostic de schizophrénie selon le DSM-III-R, les critères A, B, C, D, E, F sont monothétiques tandis que les items 1 à 5 du critère Al sont polythétiques.

Le risque d'obtenir des catégories hétérogènes avec un système diagnostique est majeur lorsque celui-ci contient plusieurs critères polythétiques. Ceci est particulièrement évident, comme l'a souligné Berner (1989), lorsqu'un tel système contient un critère polythétique comportant au choix des symptômes de premier rang de Schneider et des symptômes fondamentaux de Bleuler.

Par ailleurs, les critères n'étant pas définitifs, ils peuvent évoluer ou se modifier à l'intérieur d'un même système diagnostique. C'est ainsi que l'on a assisté à l'apparition successive des différentes classifications américaines (DSM-I, 1952; DSM-II, 1968; DSM-III, 1980; DSM-III-R, 1987; DSM-IV, 1994). Certes, les groupes de patients sélectionnés sont proches les uns des autres mais leur recouvrement n'est pas total. Fenton *et al.* (1988) en comparant le DSM-III et le DSM-III-R ont montré que ce dernier était plus restrictif en raison essentiellement des critères symptomatiques requis (critère A).

On conçoit donc que l'utilisation de tel ou tel système diagnostique pour la schizophrénie aboutisse à la sélection de groupes de patients différents. Quelques études confirment cet impressionisme clinique :

1) Les études de liaison entre deux systèmes de critères diagnostiques pris deux à deux aboutissent à une mauvaise concordance (Brockington *et al.*, 1978 ; Overall et Hollister, 1979; Stephens *et al.*, 1980; Young *et al.*, 1982; Dollfus *et al.*, 1993-a).

2) Le devenir de la maladie paraît différent selon que le patient répond initialement à tel ou tel critère diagnostique. Hwu *et al.* (1988) ont comparé le devenir sur sept ans des patients diagnostiqués initialement schizophrènes selon les critères ICD-9 et DSM-III. Ceux appartenant à l'ICD-9 ont un devenir plus favorable que ceux du groupe DSM-III en terme de déficit, de sévérité symptomatique et de fonctionnement social. De même, les patients diagnostiqués par les critères DSM-III et Feighner ont des symptômes plus durables et une détérioration socio-professionnelle plus sévère que ceux définis par les seuls critères Catego et RDC.

3) D'autre part, nous avons pu montrer les rapports entre onze catégories diagnostiques et les catégories «positive», «négative» et «dépressive» (Dollfus *et al.*, 1992-b).

Les patients de type «négatif» sont inclus préférentiellement dans certains systèmes diagnostiques tels l'ICD-9 et le système Taylor (85,7 % des inclusions) ou le système Langfeldt (80,9 % des inclusions) pour la schizophrénie. À l'inverse, les critères de Vienne incluent très peu de patients négatifs (28,6 % d'entre eux seulement) et significativement moins que les critères précédents.

Les patients «positifs» sont, eux aussi, inclus de manière très variable d'un système à l'autre : aucun patient «positif» n'est inclus dans le système Feighner, alors que l'inclusion des patients «positifs» est de 100 % pour le système New-Haven.

D'autre part, certains systèmes diagnostiques incluent préférentiellement les patients «dépressifs». Il s'agit des systèmes ICD-9 (94,4 %), New-Haven (88,8 %), Catego (77,7 %) et Langfeldt (77,7 %). Seuls les systèmes de Feighner et de Carpenter en incluent moins de la moitié (respectivement 22,7 % et 38,8 %).

4) En ce qui concerne les recherches biologiques, les divergences entre les résultats pourraient aussi découler de l'utilisation de systèmes diagnostiques différents d'une étude à l'autre. Keks *et al.* (1990) l'ont bien montré au cours d'un travail où 11 diagnostics pour la schizophrénie ont été établis rétrospectivement : les taux de prolactinémie étaient plus bas chez les patients répondant à quatre des onze catégories diagnostiques (RDC, DSM-III, Taylor-Abrams et Feighner), alors que les patients répondant aux autres systèmes avaient des taux non distincts de ceux des témoins.

Toutes ces données montrent la difficulté de comparer les résultats des études cliniques ou biologiques utilisant des systèmes diagnostiques différents et le caractère beaucoup trop partiel et hasardeux d'études qui n'utiliseraient qu'un seul système diagnostique. La meilleure stratégie pour la recherche serait donc d'utiliser, pour l'inclusion des patients, le plus grand nombre de systèmes diagnostiques reconnus afin d'aboutir à des analyses de données diversifiées et comparables. Cette approche polydiagnostique permettrait de choisir son ou ses systèmes diagnostiques selon le type d'étude envisagée :

– L'utilisation de systèmes étroits et focalisés ne serait intéressante que pour l'étude d'éventuels sous-groupes (l'ICD-9 pour l'étude de patients négatifs ou New-Haven pour l'étude de patients schizophrènes ayant une symptomatologie dépressive...).

– L'utilisation simultanée de plusieurs systèmes diagnostiques permettrait d'obtenir des populations plus représentatives de la maladie à étudier. L'évaluation diagnostique devrait être réalisée en associant des systèmes très inclusifs à d'autres qui le sont peu, ceci afin d'inclure des patients appartenant à tous les sous-groupes de la schizophrénie sans privilégier l'un d'entre eux.

C'est dans cette perspective que nous avons élaboré un outil polydiagnostique informatisé. Notre travail s'intègre dans la voie générale ouverte par plusieurs auteurs qui ont souligné l'intérêt d'une telle approche polydiagnostique pour les troubles mentaux (Berner *et al.*, 1982; Kendell, 1982; Pull, 1984; Ban, 1989).

Certains systèmes permettent d'établir tous les diagnostics catégoriels possibles (troubles anxieux, dépressifs, schizophréniques...) à l'aide d'entretiens standardisés, mais n'ont pas une approche polydiagnostique spécifique d'un trouble mental donné. Ces outils informatisés sont : le PSE-Catego (Wing *et al.*, 1974), supplanté par le SCAN (*Schedule for Clinical Assessments in Neuropsychiatry*; Wing *et al.*, 1990), le RDI (*Renard Diagnostic Interview*, Helzer *et al.*, 1981), le DIS (*Diagnostic Interview Schedule*, Robins *et al.*, 1981) et le CIDI (*Composite International Diagnostic Interview*, Robins *et al.*, 1988).

Concernant partiellement ou exclusivement la schizophrénie, des outils et checklistes sont apparus (Berner *et al.*, 1982; Endicott et Spitzer, 1978; Landmark, 1982; Overall et Hollister, 1979; Pull *et al.*, 1981; Andreasen, 1985). Plus récemment, quatre systèmes polydiagnostiques informatisés axés sur la schizophrénie, les troubles affectifs et schizoaffectifs se sont développés. Il s'agit de l'entretien standardisé (PODI) de Phillips et Maier (1986), du RPMIP (*Royal Park Multidiagnostic Instrument for Psychosis*) de McGorry *et al.* (1990), de l'OPCRIT de McGuffin *et al.* (1991) et de la LIDE (liste d'items à visée diagnostique et évolutive) élaborée dans le groupe de recherche du Professeur Petit (Dollfus *et al.*, 1994). L'intérêt de la LIDE par rapport aux trois autres outils est qu'elle peut être appliquée à des patients en phase aiguë, stabilisée voire en rémission et qu'elle inclut des critères européens. De plus, ces trois systèmes écrits en langue anglaise n'ont pas fait l'objet, jusqu'à ce jour, d'une traduction et d'une validation françaises.

MÉTHODES

Élaboration de l'outil polydiagnostique

Notre outil polydiagnostique pour la schizophrénie (LIDE) a été construit après élaboration d'une checkliste d'items et d'une informatisation du système.

Dans une première version, cette checkliste comprenait 208 items. Elle a été étendue ensuite à 241 items car nous y avons inséré les items constituant les critères de l'ICD-10 pour les troubles schizophréniques et apparentés.

Checkliste

• *Checkliste de 241 items*

Cette checkliste de 241 items, inspirée dans son principe de la «liste intégrée de critères d'évaluation taxinomiques» pour les psychoses non affectives (LICET-S) de Pull *et al.* (1981) qui comprend 70 items, a été construite après consensus entre deux psychiatres à partir des items originaux de 30 systèmes diagnostiques pour les psychoses non maniaco-dépressives et de 4 systèmes diagnostiques pour les troubles affectifs majeurs. Ceux-ci ont été pris en compte car ils constituent des critères d'exclusion pour certains systèmes pour la schizophrénie.

De même, tous les items appartenant aux formes cliniques (hébéphrénique, catatonique, paranoïde, désorganisée) de certains systèmes pour la schizophrénie (DSM-III-R, ICD-9, ICD-10, RDC) et aux formes évolutives (DSM-III-R, ICD-10, RDC) ont aussi été retenus.

Les 15 systèmes diagnostiques pour la schizophrénie ont été décrits, pour la plupart, dans les ouvrages de Landmark (1982) et de Berner *et al.* (1987) : E. Bleuler, 1911, éd. 1950; Catego S+ (Wing *et al.*, 1974); DSM-III-R, 1987; Feighner, 1972; ICD-9 (OMS, 1978); ICD-10 (OMS, 1992); Kraepelin, éd. 1919; Langfeldt, 1960-1969; RDC (Spitzer *et al.*, 1978); Schneider, 1957 – Symptômes de premier rang; index schizophrénique de New-Haven (Astrachan *et al.*, 1972); système flexible pour le diagnostic de schizophrénie (*cut-off* à 5 ou 6; Carpenter *et al.*, 1973); critères de recherche de Vienne (Berner, 1965-1969; Taylor et Abrams, 1978); critères empiriques français des schizophrénies chroniques (Pull *et al.*, 1987-a, b).

Les 15 systèmes diagnostiques utilisés pour les psychoses non maniaco-dépressives autres que la schizophrénie ont été choisis d'après les systèmes DSM-III-R, RDC, ICD-10 et les critères empiriques français de Pull *et al.* (1987-a, b) : DSM-III-R (1987) (trouble délirant; trouble schizophréniforme; psychose réactionnelle brève; trouble schizoaffectif; trouble psychotique induit); RDC (Spitzer, 1978) (trouble schizoaffectif, type maniaque; trouble schizoaffectif, type dépressif); critères empiriques français (Pull *et al.*, 1987-a, b) (psychose hallucinatoire chronique; trouble chronique interprétatif; bouffée délirante; accès délirants aigus ou subaigus considérés comme schizophréniques); ICD-10 (OMS, 1991) (trouble schizotypique; troubles délirants persistants; troubles psychotiques aigus et transitoires; troubles schizoaffectifs).

Les 4 systèmes diagnostiques pour les troubles affectifs majeurs sont le DMS-III-R (1987), le système de Feighner (1972), l'ICD-10 (OMS, 1992) et les RDC (Spitzer *et al.*, 1978).

Cette checkliste comprend des informations sociales et cliniques sur l'état actuel et antérieur de la maladie ainsi que des données familiales. Des informations sur les antécédents sont indispensables pour établir le diagnostic de certaines formes de schizophrénie, telles les formes résiduelles (DSM-III-R, RDC ou ICD-10) ou en rémission (DSM-III-R), et ont donc été prises en compte. Le recueil des données actuelles et antérieures permet ainsi de porter un diagnostic sur l'ensemble de la vie du patient ou à un moment donné.

Chaque item de la checkliste est défini dans un glossaire. La passation de la checkliste nécessite un entretien d'une moyenne de 2 heures et la cotation des items nécessite 30 minutes à 2 heures selon l'importance des antécédents du sujet.

• *Checkliste de 183 items*

Une version plus réduite (183 items) a été extraite de la précédente dans la perspective d'un travail axé uniquement sur la schizophrénie et la psychose hallucinatoire chronique (PHC). Cette checkliste comprend les items correspondant à 14 systèmes diagnostiques pour la schizophrénie, au diagnostic de la psychose hallucinatoire chronique (PHC) définie par Pull *et al.* (1987-a, b) et aux critères d'exclusion sus-décrits. La PHC a été sélectionnée dans la mesure où elle pouvait être étroitement liée à la schizophrénie (Dollfus *et al.*, 1992a).

Programme informatisé

À partir de cette checkliste, les différents diagnostics ont été reconstruits avec un système de codage après consensus entre deux psychiatres. Ce codage a ensuite été retranscrit sous forme d'algorithmes de classement et programmé en langage GW-Basic après consensus entre un informaticien et un psychiatre. Ainsi, un tel système permet d'attribuer à un patient donné l'un de ces multiples diagnostics de manière instantanée alors que le cotateur peut rester aveugle par rapport à toutes répartitions diagnostiques.

Validation de l'outil polydiagnostique

Une fidélité inter-juges de la checkliste (LIDE) et une validation du programme informatique lui-même ont été pratiquées sur 15 patients.

Les critères d'inclusion étaient initialement très larges de façon à obtenir une population hétérogène : les patients devaient présenter des hallucinations et/ou des idées délirantes et/ou des symptômes négatifs tels un émoussement affectif, un repli ou un comportement désorganisé et ne devaient pas répondre aux troubles dépressif ou maniaque majeurs (DSM-III-R), ni présenter des symptômes secondaires à un trouble organique.

Fidélité inter-juges

La fidélité inter-juges a été appréciée selon la méthode de « l'examinateur-l'observateur », ce qui a nécessité au préalable l'entraînement d'un cotateur par l'investigateur principal avec dix examens « à blanc ».

Les 15 patients ont ensuite été examinés simultanément par les deux cotateurs. L'entretien a été mené successivement (après randomisation) par l'un des deux. Ces

entretiens n'étaient pas structurés; toutefois, les données recueillies sur la LIDE étaient parfaitement standardisées. Bien sûr, ces deux examinateurs ont rempli indépendamment la checkliste.

Validation du programme

La validité du programme informatique a consisté à examiner la cohérence existant entre les diagnostics établis par un autre clinicien et les résultats diagnostiques obtenus par le programme informatique et ce à partir d'un même recueil de données (la checkliste).

Les 15 checklistes précédentes, remplies par un cotateur et pour lesquelles un (ou plusieurs) diagnostic(s) a (ont) été établi(s) par le système informatisé, ont fait l'objet d'un examen par un autre clinicien, à l'insu du patient et de toute répartition diagnostique. Ce cotateur, après familiarisation avec la LIDE, a réparti manuellement les patients dans une ou plusieurs de ces 15 catégories diagnostiques, à partir des items des checklistes et à l'aide visuelle des critères opérationnels.

Sept hommes (46,7 %) et 8 femmes (53,3 %) ont été inclus. Onze patients (73,3 %) étaient, lors de la cotation, en phase aiguë (premier épisode ou exacerbation de la maladie). L'âge moyen était de 33,6 ± 12,9 ans et la durée de la maladie de 7 ± 10,5 années.

Les coefficients kappa de Cohen obtenus entre les deux paires de cotateurs pour chaque système diagnostique sont représentés sur le tableau 1.

Ce tableau montre que la concordance entre les deux examinateurs, permettant de tester la validation du programme (test B), est toujours égale ou supérieure à celle appréciant la fidélité inter-juges. Le test de validation du programme lui-même montre que les concordances sont toutes excellentes (kappa de 0,75 à 1). Neuf systèmes diagnostiques obtiennent un kappa à 1. Pour les six autres systèmes où le kappa oscille entre 0,76 et 0,89, des explications ont été recherchées et trouvées ne permettant dans aucun cas d'incriminer le programme informatique. Ces résultats sur la validité du programme informatique sont satisfaisants si l'on se réfère aux données de Spitzer *et al.* (1974) sur le programme informatique DIAGNO. Notre méthode informatique s'écarte, toutefois, de la leur ainsi que d'autres basées sur la théorie des probabilités. L'utilisation de telles méthodes a comme conséquence d'établir les diagnostics avec une certaine probabilité et donc avec un risque accru de faux positifs ou de faux négatifs. Notre procédure est plus exacte dans la mesure où nous avons retranscrit exactement les critères opérationnels en format ordinateur. Si l'avantage d'une telle procédure est une lecture informatique exacte du critère diagnostique, la durée de la passation et de la saisie des données la fait réserver à la recherche.

Pour ce qui est de la fidélité inter-juges (test A), les deux cotateurs ont une concordance excellente (kappa de 0,75 à 1) pour les diagnostics de schizophrénie selon les systèmes Catego, DSM-III-R, Feighner, ICD-9, Langfeldt, Pull, Schneider et Taylor-Abrams. Ces deux mêmes cotateurs ont une bonne concordance diagnostique pour les 6 autres systèmes (kappa de 0,40 à 0,75) et pour le diagnostic de PHC (kappa = 0,71). Ces résultats sont tout à fait compatibles avec ceux obtenus avec d'autres procédures diagnostiques. Certes, l'existence d'un entretien semi-structuré permettrait vraisemblablement d'améliorer encore la fidélité inter-juges; il est cependant à noter que des systèmes comme le DIAGNO (Spitzer *et al.*, 1974) ne possèdent pas, sans atteinte à leur fiabilité, d'entretiens semi-structurés.

Tableau 1. *Coefficients kappa pour chaque catégorie diagnostique (diagnostic probable ou certain)*

	Fidélité inter-juges (test A)		*Validation du programme (test B)*	
	K	*DS*	*K*	*DS*
1) Troubles schizophréniques				
E. Bleuler	0,52**	0,23	0,88****	0,22
Carpenter *cut-off* à 6	0,52	0,30	0,86***	0,28
Catego	1***	0,29	1***	0,28
DSM-III-R (forme actuelle et résiduelle)	0,86***	0,26	0,86***	0,26
Feighner (forme actuelle et résiduelle)	1**	0,14	1**	0,37
ICD-9	1	0,68	1	0,68
ICD-10	0,70**	0,29	1****	0,26
Langfeldt	1***	0,29	1***	0,28
New-Haven	0,58	0,37	0,76*	0,41
Pull	0,86***	0,27	1****	0,26
RDC (forme actuelle et résiduelle)	0,59***	0,19	0,89****	0,19
Schneider	0,86***	0,27	1***	0,32
Taylor-Abrams	0,84**	0,30	0,84**	0,30
Vienne	0,68***	0,19	1****	0,19
2) Troubles non schizophréniques				
Psychose hallucinatoire chronique (Pull)	0,71**	0,27	1****	0,26

k = kappa de Cohen; DS : déviation standard du kappa; * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$; **** $p < 0,0001$.

Une telle approche polydiagnostique paraît actuellement indispensable pour la recherche. En effet, elle seule permet l'étude de la schizophrénie au sens large du terme sans privilégier un sous-groupe particulier comme le ferait un système diagnostique utilisé seul. À l'opposé, elle permet également l'étude focalisée et comparative des patients entrant dans chaque système diagnostique. C'est dans cette perspective que deux études ont été réalisées :

– La première a consisté à établir les liens entre la psychose hallucinatoire chronique et la schizophrénie, définie non pas par un seul système diagnostique mais par plusieurs systèmes (Dollfus *et al.*, 1992-a). Nous avons pu ainsi montrer une liaison significative entre la PHC et le diagnostic de schizophrénie selon les critères Catego-S+, New-Haven et Schneider mais à l'inverse une parfaite indépendance entre la PHC et le diagnostic de schizophrénie selon les critères de Bleuler.

– La deuxième étude (Dollfus *et al.*, 1993-b) a permis de montrer des variations considérables du nombre de patients inclus d'un système à l'autre et d'une phase de la maladie à une autre (phases aiguë, post-aiguë, stable). Les systèmes qui ne comprennent que des critères renfermant des symptômes positifs se montrent particulièrement instables. À l'inverse, d'autres critères stabilisent les systèmes; il s'agit, certes, du critère de durée de la maladie, des critères d'adaptation sociale mais aussi

des symptômes négatifs tels l'émoussement affectif ou le repli social. La seule présence de ces symptômes négatifs suffirait ainsi à stabiliser un système et permettrait même de considérer comme redondant le critère temps qui oblige à porter un diagnostic de schizophrénie après un délai de 6 mois (pour les critères de Feighner et du DSM-III-R) voire de 5 ans pour les critères de Langfeldt.

Cette approche polydiagnostique mériterait d'être davantage considérée pour les recherches sur la schizophrénie, en particulier génétiques, biochimiques et neuroanatomiques.

Ainsi, certains travaux préliminaires comme celui de Vasse (1993) soulignent cette nécessité de tester des associations entre des gènes codant pour les récepteurs dopaminergiques et des groupes de patients définis par des systèmes de critères diagnostiques différents. Dans la mesure où il existe probablement plusieurs schizophrénies, comme d'ailleurs l'évoquait Bleuler au début du siècle, le choix d'un seul système diagnostique, comme le DSM-III-R, s'avère être encore totalement arbitraire.

Bibliographie

Andreasen N. C. — *Comprehensive Assessment of Symptoms and History (CASH).* IA : the university of Iowa, Iowa city, 1985.

Ban T. A. — *CODE-DD Composite Diagnostic Evaluation of Depressive Disorders.* Code Distributors, Nashville, 1989.

Berner P. — À propos de la méthodologie du suivi des psychoses au long cours. In *Colloque des psychoses au long cours.* Sous la direction de J. Guyotat. Specia éd., Paris, 1989.

Berner P., Gabriel E., Katschnig H., Kieffer W., Koehler K., Lenz G., Simhandl C. H., Pull C. B., Pull M. C., Berner C. — *Critères diagnostiques pour les psychoses schizophréniques et affectives.* Expansion scientifique française, Paris, 1987.

Berner P., Katschnig H., Lenz G. — Polydiagnostic Approach : a Method to Clarify Incongruences among the Classification of the Functional Psychoses. *Psychiatr. J. Univ., 7,* 244-248, 1982.

Brockington I. F., Kendell R. E., Leff J. P. — Definitions of Schizophrenia : Concordance and Prediction of Outcome. *Psychological Medecine, 8,* 387-398, 1978.

Diagnostic and Statistical Manual of Mental Disorders. Third edition, revised (DMS-III-R). American Psychiatric Association, Washington DC. 1987. Traduction française : *Manuel diagnostique et statistique des troubles mentaux.* Masson, Paris, 1989.

Dollfus S., Petit M., Ménard J. F., Brazo P., Besse-Assouly F., Preterre Ph., Lesieur Ph. — *Polydiagnostic Approach of Schizophrenia : Validity of a Computer Checklist (LIDE). L'Encéphale,* XX, 91-101, 1994.

Dollfus S., Petit M., Ménard J. F. — The Relationship between « Chronic Hallucinatory Psychosis » (CHP) and Schizophrenia. *Eur. Psychiatry,* 7, 271-276, 1992-a.

Dollfus S., Petit M., Ménard J. F., Lesieur Ph. — Recherche sur la schizophrénie : nécessité d'inclure les patients selon plusieurs systèmes diagnostiques. *Ann. Méd. Psychol., 150,* n° 4-5, 327-331, 1992-b.

Dollfus S., Petit M., Ménard J. F., Lesieur Ph. — Schizophrenia : Comparison of 13 Diagnostic Systems in a Cross-Sectional Study. *Eur. Psychiatry, 8,* 7-13, 1993-a.

Dollfus S., Petit M., Ménard J. F., Lesieur Ph. — Symptomatic and Diagnostic Stability in Schizophrenia : a One-Year Follow-up Study. *Eur. Psychiatry, 8,* 131-138, 1993-b.

Endicott J., Spitzer R. L. — A Diagnostic Interview : the Schedule for Affective Disorders and Schizophrenia. *Arch. Gen. Psychiatry, 35,* 837-844, 1978.

FEIGHNER J. R., ROBINS E., GUZE S. B., WOODRUFF R.A., WINOKUR G., SIMON R. — Diagnostic Criteria for Use in Psychiatric Research, *Arch. Gen. Psychiatry*, *26*, 57-63, 1972.

FENTON W. S., MCGLASHAN T. H., HEINSSEN R. K. — A Comparison of DSM-III and DSM-III-R Schizophrenia. *Am. J. Psychiatry*, *145*, 1446-1449, 1988.

HELZER J. E., ROBINS L. N., CROUGHAN J. L., WELNER A. — Renard Diagnostic Interview. Its Reliability and Procedural Validity with Physicians and Lay Interviewers. *Arch. Gen. Psychiatry*, *38*, 393-398, 1981.

HWU H. G., CHEN C. C., STRAUSS J. S., TAN K. L., TSUANG M. T., TSENG W. S. — A Comparative Study on Schizophrenia Diagnosed by ICD-9 and DSM-III : Course, Family History and Stability of Diagnosis. *Acta Psychiatr. Scand.*, *77*, 87-97, 1988.

KEKS N. A., COPOLOV D. L., KULKARNI J., MACKIE B., SINGH B., MCGORRY P., RUBIN R. T., HASSETT A., MCLAUGHLIN M., VAN RIEL R. — Basal and Haloperidol-Stimulated Prolactin in Neuroleptic-free Men with Schizophrenia Defined by 11 Diagnostic Systems. *Biol. Psychiatry*, *27*, 1203-1215, 1990.

KENDELL R. E. — The Choice of Diagnostic Criteria for Biological Research. *Arch. Gen. Psychiatry*, *39*, 1334-1339, 1982.

LANDMARK J. — A Manual for the Assessment of Schizophrenia. *Acta Psychiatr Scand*, *65 (Suppl. 298)*, 1982.

MCGORRY P., SINGH B., COPOLOV D. L., KAPLAN I., DOSSETOR C. R., VAN RIEL R.J. — Royal Park Multidiagnostic Instrument for Psychosis : Part II. Development, Reliability, and Validity. *Schizophrenia Bull.*, *16*, 517-536, 1990.

MCGUFFIN P., FARMER A. E., HARVEY I. — A Polydiagnostic Application of Operational Criteria in Studies of Psychotic Illness : Development and Reliability of the OPCRIT System. *Arch. Gen. Psychiatry*, *48*, 764-770, 1991.

MOREL B. A. — *Traités des maladies mentales*. Masson, Paris, 1860.

ORGANISATION MONDIALE DE LA SANTÉ — CIM-10. *Classification internationalee des maladies. Dixième révision. Chapitre V (F) : troubles mentaux et troubles du comportement.* Masson, Paris, 1992.

OVERALL J. E., HOLLISTER L. E. — Comparative Evaluation of Research Diagnostic Criteria for Schizophrenia. *Arch. Gen. Psychiatry*, *36*, 1198-1205, 1979.

PHILIPPS M., MAIER W. — The Polydiagnostic Interview : a Structured Interview for the Polydiagnostic Classification of Psychiatric Patients. *Psychopathology*, *19*, 175-185, 1986.

PULL C. B., PULL M. C., PICHOT P. — LICET-S : une liste intégrée de critères d'évaluation taxinomiques pour les psychoses non affectives. *J. Psy. Biol. Therap.*, *1*, 1, 33-37, 1981.

PULL C. B., PULL M. C., PICHOT P. — Des critères empiriques français pour les psychoses. I. Position du problème et méthodologie. *Encéphale*, *X*, 119-123, 1984.

PULL M. C., PULL C. B., PICHOT P. — Des critères empiriques français pour les psychoses. II. Consensus des psychiatres français et définitions provisoires. *Encéphale*, *XIII*, 53-57, 1987-a.

PULL M. C., PULL C. B., PICHOT P. — Des critères empiriques français pour les psychoses. III. Algorithmes et arbre de décision. *Encéphale*, *XIII*, 59-66, 1987-b.

ROBINS L. N., HELZER J. E., CROUGHAN J., RATCLIFF K.S. — National Institute of Mental Health Diagnostic Interview Schedule. Its History, Characteristics, and Validity. *Arch. Gen. Psychiatry*, *38*, 381-389, 1981.

ROBINS L. N., WING J, WITTCHEN H. U., HELZER J. E., BABOR T. F., BURKE J., FARMER A., JABLENSKI A., PICKENS R., REGIER D. A., SARTORIUS N., TOWLE L. H. — The Composite International Diagnostic Interview. *Arch. Gen. Psychiatry*, *45*, 1069-1077, 1988.

SPITZER R. L., ENDICOTT J., COHEN J., FLEISS J. L. — Constraints on the Validity of Computer Diagnosis. *Arch. Gen. Psychiatry*, *31*, 197-203, 1974.

SPITZER R. L., ENDICOTT J., ROBINS E. — Research Diagnostic Criteria. *Arch. Gen. Psychiatry*, *35*, 773-782, 1978.

STEPHENS J. H., OTA K. Y., CARPENTER W. T., SHAFFER J. W. — Diagnosis Criteria for Schizophrenia : Pronostic Implications and Diagnostic Overlap. *Psychiatry Research*, *2*, 1-12, 1980.

VASSE T. — Schizophrénie et génétique. Utilisation d'un outil polydiagnostique, la LIDE dans une étude d'association. Mémoire DES de psychiatrie. Rouen, avril 1993.

VAUGHN C., LEFF J., WING J. K., COOPER J. E., SARTORIUS N. — The Measurement and Classification of Psychiatric Symptoms. Cambridge University Press, London, 1974.

WING J. K., BABOR T., BRUGHA T., BURKE J., COOPER J. E., GIEL R., JABLENSKI A., REGIER D., SARTORIUS N. — SCAN : Schedules for Clinical Assessment in Neuropsychiatry. *Arch. Gen. Psychiatry*, *47*, 6, 589-594, 1990.

YOUNG M. A., TANNER M. A., MELTZER H. Y. — Operational Definitions of Schizophrenia. What do they identify? *J. Nervous and Mental Disease*, *170*, 443-447, 1982.

4

PROBLÈMES POSÉS PAR LE CHOIX D'UNE ÉCHELLE

par J. FERMANIAN

La construction puis la validation d'une nouvelle échelle nécessitent de grands moyens matériels et humains. Ce travail de longue haleine s'effectue en plusieurs étapes successives permettant d'améliorer progressivement l'instrument de départ; au total, il faut généralement plusieurs années pour aboutir à une version ayant des qualités métrologiques satisfaisantes.

Plutôt que se lancer dans l'aventure d'une nouvelle construction, tous les auteurs recommandent donc d'utiliser une des nombreuses échelles déjà disponibles. Pour effectuer un choix judicieux, une méthodologie très rigoureuse basée sur une comparaison approfondie des qualités métrologiques de chaque instrument est absolument nécessaire.

Les trois qualités fondamentales sont validité, reproductibilité (*reliability*), structure factorielle.

Remarque importante :

Une certaine confusion existe dans la terminologie internationale. Certains auteurs nomment différemment le même concept; la traduction française d'un terme anglais peut être variable, etc. Afin d'éviter les ambiguïtés, nous utiliserons donc le vocabulaire défini par l'*American Psychological Association*.

VALIDITY
(Nunnally, 1970, 1978)

Une échelle est dite valide si elle mesure correctement le phénomène qu'elle est censée mesurer. Ce concept très général se décompose en fait en différentes facettes.

Face validity

Pour l'évaluer, un juge compétent après avoir inspecté très attentivement l'instrument se prononce sur sa valeur apparente; elle lui semble satisfaisante ou non. La *face validity* résulte donc d'un jugement subjectif, fonction de l'observateur, prenant en compte les aspects visibles de l'échelle : longueur, libellé des items, modalités de réponse, etc.

Trop souvent, on rejette un instrument en se basant uniquement sur sa *face validity*, certains items par exemple ne semblent pas totalement satisfaisants.

Content validity

Elle résulte également de l'appréciation d'un juge compétent. Ce dernier étudie chaque item afin de voir s'il explore au moins une dimension du phénomène mesuré. Ceci nécessite donc d'avoir établi au préalable la liste des dimensions importantes du phénomène.

Par exemple, le ralentissement chez les déprimés est constitué principalement de deux dimensions : motrice et idéique. La *content validity* de l'échelle ERD, mesurant ce ralentissement, peut être considérée comme satisfaisante car de nombreux items investiguent la dimension motrice et tous les autres explorent la dimension idéique. Le nombre d'items correspondant à chaque dimension dépend généralement de l'importance relative de chacune des dimensions dans la structure du phénomène mesuré. Ainsi, l'ERD contient plus d'items moteurs que d'items idéiques.

Notons que la *face validity* et la *content validity* résultent toutes deux du jugement subjectif d'une personne compétente, c'est pourquoi certains auteurs réunissent ces deux types de validité en une seule entité commune.

Criterion validity

Pour mesurer cette validité, un critère extérieur à l'échelle évaluant le même phénomène est pris comme référence ou « gold standard ». Un nombre suffisant de sujets sont alors évalués à la fois par l'échelle et le critère de référence, et l'on mesure l'intensité du lien statistique existant entre les deux évaluations. Le « gold standard » peut être l'opinion d'un ou de plusieurs experts ou bien le score obtenu à une autre échelle considérée par toute la communauté scientifique comme une référence indiscutable.

On distingue deux types de *criterion validity*.

Concurrent validity

Le critère de référence et l'échelle sont utilisés en même temps chez tous les sujets ; en général, au cours du même entretien.

Predictive validity

Le phénomène est d'abord mesuré par l'échelle étudiée, puis après un certain temps seulement par le critère de référence. Par exemple, dans le cas d'une échelle destinée à prédire le pronostic, la référence prise habituellement est le pronostic réel des sujets à plus ou moins longue échéance.

Au cours de l'évaluation de la *criterion validity*, deux points sont particulièrement importants :

a) Le choix du critère de référence qui doit être accepté par tous comme un « gold standard » incontestable.

b) Le méthodes statistiques utilisées pour mesurer le lien entre les deux évaluations :

– Soit de simples coefficients de corrélation, paramétrique ou non paramétrique selon que les variables utilisées sont quantitatives ou ordonnées.

– Soit de véritables coefficients d'association dans un tableau de contingence, et non un simple test du Khi-deux, lorsque les deux variables sont qualitatives.

Construct validity

Deux étapes

Elles sont nécessaires pour mesurer correctement la *construct validity* :

a) On postule d'abord qu'il existe, en raison de considérations théoriques, un lien entre :

– le phénomène mesuré, par exemple la mémoire ;

– un certain indicateur, par exemple le temps mis à accomplir une tâche donnée.

b) Ensuite, on met en place une expérimentation sur plusieurs sujets afin de mesurer l'intensité du lien existant réellement entre :

– le phénomène mesuré, évalué au moyen de l'échelle de mémoire étudiée ;

– et l'indicateur.

Si l'intensité observée du lien est forte, cette expérimentation apporte un argument en faveur de la *construct validity* de l'échelle.

Celle-ci ne peut être déduite des résultats d'une étude unique, il faut effectuer plusieurs expérimentations avec divers indicateurs avant de conclure à la *construct validity* de l'échelle.

Convergent et discriminant validity

Ce sont deux formes particulièrement importantes de *construct validity*.

a) Si dans l'exemple de l'échelle de mémoire étudiée ci-dessus, on observe chez les sujets de l'expérimentation une **forte corrélation** entre le score de l'échelle et le temps mis à accomplir la tâche comme cela est postulé en théorie, alors, on a un argument en faveur de la *convergent validity*.

b) Si au cours d'une autre expérimentation, on ne constate **pas de corrélation** entre l'intelligence et la mémoire et que l'on a postulé l'indépendance de ces deux fonctions dans une étape préalable, ceci est alors un argument en faveur de la *discriminant validity*.

Méthode des groupes extrêmes

Cette technique de validation est proche de la recherche de la *convergent validity*. On constitue deux groupes de sujets nettement différents pour le phénomène étudié, par exemple : cliniquement déprimés et non déprimés. Les moyennes des scores obtenus dans chacun de ces deux groupes à une échelle de dépression doivent alors être significativement différentes.

La méthode multitrait-multimethod matrix

Elle permet d'étudier simultanément la *convergent* et la *discriminant validity*. Nous ne l'exposerons pas ici.

Remarques

a) La recherche de la *construct validity* est un processus continu, nécessitant plusieurs expérimentations. Cette validité est confortée peu à peu au fil des études successives.

b) On vérifie au cours d'une même expérimentation à la fois des considérations théoriques et leurs conséquences.

RELIABILITY

Une échelle possède cette qualité si, utilisée pour mesurer plusieurs fois le même phénomène, dans les mêmes conditions, elle donne des résultats voisins. En pratique, on distingue plusieurs types de *reliability*.

Inter observer reliability (Fermanian, 1984, 1989)

Deux observateurs cotent de manière indépendante N sujets en utilisant la même échelle ; à partir de ces cotations, on mesure l'accord observé entre les deux juges.

Si l'on s'intéresse à l'accord obtenu sur le score global de l'échelle qui est une donnée quantitative, on peut résumer l'ensemble des résultats au moyen d'un tableau à N lignes. Chaque ligne correspond à un sujet, et comprend les deux scores attribués par chaque observateur à ce sujet. On est alors dans le cas quantitatif (Fermanian, 1984).

On peut aussi résumer l'accord des deux juges sur leurs réponses à un seul item. Si cet item est coté par exemple en quatre modalités ; 0, +, ++, +++, les résultats sont alors résumés dans un tableau carré à $4 \times 4 = 16$ cases, dit tableau de contingence des effectifs. Chaque case contient le nombre de sujets correspondant au croisement des cotations des deux juges. Par exemple, si la case 0, ++ contient l'effectif 9, cela signifie que 9 sujets ont été cotés à la fois 0 par le premier observateur et ++ par le deuxième. On se trouve dans le cas qualitatif (Fermanian, 1984).

Intra observer reliability

Un seul observateur cote tous les sujets à deux temps différents. En admettant que le phénomène mesuré reste stable entre les deux temps, on mesure alors l'accord observé entre les deux cotations.

Comme précédemment si l'on analyse le score global de l'échelle, on se trouve dans le cas quantitatif, et si l'on s'intéresse à l'accord sur les modalités de réponse d'un seul item, on est dans le cas qualitatif.

Test-retest reliability

Le cas classique est celui d'un autoquestionnaire où le sujet cote lui-même son état; il n'y a donc pas d'observateur externe. Comme ci-dessus, on lui demande d'effectuer ses cotations à deux temps différents et on admet que son état reste stable entre ces deux mesures.

Là encore, on distingue le cas quantitatif : mesure de l'accord sur le score global, du cas qualitatif : mesure de l'accord sur les réponses à un item.

Mesures de l'internal consistency

Différentes méthodes évaluant la cohérence interne de l'échelle ou *internal consistency* sont souvent utilisées pour mesurer la *reliability*.

Item-total corrélation

On mesure la valeur du coefficient de corrélation entre chaque item et le score global diminué de la note de l'item étudié.

Split-half reliability

En tirant au sort les items, on divise l'échelle de départ en deux sous-échelles qui sont ensuite corrélées entre elles.

Coefficient α de Cronbach (Campbell et Fiske, 1959)

Il se calcule selon la formule suivante lorsque chaque item a des modalités de réponse quantitatives :

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\Sigma\sigma_i^2}{\sigma_T^2}\right)$$

avec n = nombre d'items de l'échelle,
σ_T^2 = variance du score total,
σ_i^2 = variance de chaque item.

Notons que lorsque les modalités de réponse de chaque item sont binaires, il faut utiliser la formule de Kuder-Richardson.

Attention :

– On ne peut calculer la cohérence interne que si l'on est certain que l'échelle est réellement unidimensionnelle.

– Même dans ce cas, mesurer la *reliability* au moyen de la cohérence interne est une démarche critiquable. On ne peut en effet superposer ces deux concepts, même si cela est assez souvent effectué en pratique courante.

Les techniques statistiques

Selon que l'on se trouve dans le cas quantitatif ou qualitatif, les méthodes à utiliser pour mesurer les divers types de *reliability* sont différentes.

Cas quantitatif (Fermanian, 1984b)

On ne doit pas employer le coefficient de corrélation usuel mais les coefficients de *corrélation intraclasse.* En effet, le premier coefficient mesure uniquement l'intensité de la *liaison* existante entre les notes de deux juges par exemple, alors que l'on s'intéresse à leur *accord.* Les notes doivent non seulement être corrélées mais concordantes.

Cas qualitatif (Fermanian, 1984a)

Utiliser le Khi 2 est une erreur ; en effet, on s'intéresse alors à la liaison et non à la concordance.

La reliability est correctement mesurée au moyen des coefficients *Kappa* et *Kappa pondéré* dont le calcul pratique est illustré dans la référence.

Remarques

a) Il faut un nombre de sujets N suffisant pour évaluer avec une certaine précision les coefficients précédents :

- $N \geq 40$ pour les coefficients intraclasse.
- $N \geq 30$ pour Kappa ou Kappa pondéré lorsque le nombre de modalités de réponses par item est inférieur ou égal à 4.

b) Il existe plusieurs variantes des coefficients précédents, ainsi :

- Le calcul du coefficient intraclasse diffère selon que les deux juges sont fixes ou aléatoires.
- Le système de pondération de Kappa pondéré peut être linéaire ou quadratique.

c) Les considérations précédentes peuvent être généralisées à plus de deux juges et à plus de deux cotations par sujet (Fermanian, 1989).

STRUCTURE FACTORIELLE
(Harman, 1976)

Analyse de la structure factorielle

Avant d'utiliser une échelle, il est très important de savoir si une analyse sérieuse et détaillée de sa structure factorielle a été effectuée. Si oui, une première structure est proposée généralement par le « père » de l'instrument, puis elle est confirmée ou non par des analyses ultérieures sur divers échantillons de sujets.

Analyse des résultats obtenus

Afin de pouvoir juger correctement les résultats obtenus à partir d'une étude donnée, on doit posséder les informations suivantes :

– Description précise des *sujets analysés* avec les critères d'inclusion et d'exclusion.

– Le *nombre de sujets* intervenant dans les analyses factorielles. Un minimum de 200 est souhaitable pour assurer un minimum de stabilité des résultats.

– La démarche est-elle exploratoire ou au contraire vise-t-elle à confirmer une structure donnée?

– Les *techniques statistiques* utilisées doivent être précisées et justifiées, par exemple analyse en composantes principales, seule ou avec rotation, varimax ou autre méthode, analyse des correspondances, etc.

– La description des *problèmes rencontrés* au cours de l'analyse : données manquantes, outliers...

– Les résultats obtenus seront comparés à ceux des *études précédentes.*

CONCLUSION

Pour choisir judicieusement une échelle parmi toutes celles qui sont disponibles sur le marché, il faut prendre en compte essentiellement la *validity*, la *reliability* et la structure factorielle de chacune d'elles.

Ces trois caractéristiques doivent être prises en compte *simultanément*, en effet un instrument peut avoir par exemple une bonne *reliability* mais une faible *validity.*

Enfin, cela suppose que l'on possède des informations détaillées sur les méthodes utilisées au cours des diverses validations : nature des coefficients, valeur numérique, taille des échantillons, etc. Certes, il n'est pas toujours aisé en pratique de réunir toutes ces informations mais sans elles un choix rigoureux est impossible.

BIBLIOGRAPHIE

CAMPBELL D. T., FISKE D. W. — Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin, 56,* 81-105, 1959.

CRONBACH L. J. — Coefficient Alpha and Internal Structure of Tests. *Psychometrika, 16,* 297-334, 1951.

FERMANIAN J. — Mesure de l'accord entre deux juges : cas qualitatif. *Rev. Epidém. et Santé Publ., 32,* 140-147, 1984a.

FERMANIAN J. — Mesure de l'accord entre deux juges : cas quantitatif. *Rev. Epidém. et Santé Publ., 32,* 408-413, 1984b.

FERMANIAN J. — Methodology and Statistical Problems in the Assessment of Memory by Rating Scales. *Archives of Gerontology and Geriatrics, Supp. 1,* 215-221, 1989.

HARMAN H. H. — *Modern Factor Analysis* (3e éd.), University of Chicago Press, Chicago, 1976.

NUNNALLY J. C. — *Introduction to Psychological Measurement.* McGraw-Hill, New York, 1970.

NUNNALLY J. C. — *Psychometric Theory* (2[e] éd.), McGraw-Hill, New York, 1978.

5

BATTERIES D'ÉVALUATION INFORMATISÉES

par L. HUGONOT-DIENER

DÉFINITIONS DES BATTERIES D'ÉVALUATION COGNITIVES INFORMATISÉES

L'utilisation de l'informatique a permis depuis plusieurs années l'élaboration de tests simples comme les temps de réaction à l'apparition d'un signal.

Les progrès technologiques en matière d'ordinateur et de programmes informatiques ont permis la mise au point d'évaluations psychométriques plus élaborées, en complément ou en remplacement de l'approche traditionnelle « papier-crayon ».

Ces techniques sont précieuses :

– afin de mesurer l'intelligence, l'attention et la mémoire à court terme, ainsi que l'aptitude de sujets sains par exemple pour l'orientation professionnelle ;

– en neuropsychologie expérimentale pour quantifier les théories cognitives et mieux comprendre certaines affections (en particulier celles qui résultent de lésions dégénératives du cerveau affectant les fonctions supérieures) ;

– en clinique, pour mieux évaluer le stade et l'évolution de la maladie d'un patient ;

– enfin, en pharmacoclinique pour mesurer l'effet des médicaments sur les fonctions supérieures.

Chronologiquement, sont apparus trois types de batteries informatisées.

Informatisation de la cotation et de l'interprétation de tests classiques

Certaines batteries sont des systèmes-expert pour les calculs et l'exploitation des résultats de test « papier-crayon » : type MMPI informatisé ou mini-MULT, et permettent une économie de temps et une plus grande fidélité dans la correction et l'interprétation des tests.

Informatisation des tests « papier-crayon », avec chronométrie

Certaines batteries de première génération correspondent à l'automatisation de tests existants et validés en version traditionnelle « papier-crayon », et prenant en compte la chronométrie et le temps de réaction pour toutes les réponses (comme la batterie du *Cognitive Drug Research* (CDR) ou les tests de Fondarai) ainsi que la stratégie des sujets, certains paramètres enregistrés pour tous les tests permettant

cette approche : le nombre d'erreurs, le temps de latence par réponse et le temps total par exemple.

La prise en compte du temps constitue un progrès si on compare un sujet par rapport à lui-même, mais ne peut être confondue avec l'évaluation d'un processus cognitif, car on sait bien que la lenteur d'idéation n'est pas forcément synonyme d'inefficacité.

Batteries de tests développées à partir de l'approche cognitiviste

Ces tests évaluent la durée des opérations mentales nécessaires à la réalisation d'une tâche. Comme le dit J. Poitrenaud : « les recherches basées sur les modèles de la psychologie cognitive ont clairement montré que toute tâche cognitive faisait intervenir plusieurs opérations mentales distinctes, ces opérations peuvent se dérouler simultanément dans le cas d'une tâche simple et automatisée. Lorsque la tâche est complexe, exigeant de ce fait davantage de ressources attentionnelles, le traitement s'effectue séquentiellement. Par exemple, J. Theios a proposé de distinguer cinq stades successifs de traitement de l'information dans la tâche d'attention sélective que représente le test de temps de réaction au choix. Il s'agit des étapes suivantes : intégration sensorielle des qualités physiques du stimulus; identification de ce stimulus (ou catégorisation); décision; programmation de la réponse motrice; et la réponse motrice elle-même ».

CRITÈRES DE CHOIX

Le choix d'une batterie d'examens informatisés est le même que pour toute batterie neuropsychologique.

Pour contribuer efficacement à une description de la personnalité ou à la clinique selon un aspect qualitatif et quantitatif, il doit répondre aux objectifs suivants :

– tenir compte de la validité de la batterie pour la mesure recherchée, et de l'étalonnage sur la population cible;

– contenir des épreuves standardisées tenant compte de l'influence de l'âge et du niveau culturel, influençant différemment les résultats de ces tests (Micas *et al.*, 1992);

– avoir des épreuves suffisamment variées pour mettre en évidence des variations ou des déficits parfois très sélectifs;

– comprendre des épreuves de complexité graduée pour mettre en évidence des variations légères ou des déficits éventuels chez des patients ayant une atteinte même légère des fonctions supérieures et ne pas excéder les capacités de coopération des sujets ou des patients ayant à les subir;

– permettre d'évaluer les troubles émotionnels, les plaintes fonctionnelles et le retentissement quotidien;

– enfin, bien sûr, être intégré dans l'histoire clinique du patient et l'ensemble des examens complémentaires.

AVANTAGES ET INCONVÉNIENTS PAR RAPPORT AUX TESTS TRADITIONNELS

Les avantages généraux de l'évaluation informatisée

Ce sont surtout des avantages méthodologiques :

– la standardisation des conditions de passation du test et des méthodes : avec un affichage des consignes standardisées ;

– la garantie de procédures parfaitement normalisées, systématiques, similaires : avec une uniformité de passation d'un examen à l'autre et d'un patient à l'autre : plaçant le sujet dans une situation hautement reproductible, grâce à une gestion automatisée de la passation ;

– le traitement automatisé des résultats et leur stockage ;

– un affichage des résultats et un listing individualisé de l'évaluation ;

– la prise en compte de la chronométrie, pour tous les tests de la batterie.

Certains ont montré que les tests informatisés (Hibert *et al.*, 1994) par rapport aux tests classiques, prenant en compte non seulement la performance mais aussi le temps de passation étaient parfois plus sensibles. C'est ce facteur temps qui paraît plus sensible que la performance et qui peut être un facteur discriminant précoce (Micas *et al.*, 1992) d'une atteinte d'un sujet ; ceci étant valable pour certains tests seulement comme le temps de réaction, la mémoire de travail, la décision lexicale, alors que cela n'a aucun intérêt pour la mémoire à long terme par exemple.

L'analyse des processus mentaux

Certains se sont demandés si la rapidité était un bon indicateur du fonctionnement du cerveau, et ont préféré mesurer le temps de traitement de l'information (Pater *et al.*, à paraître), ceci a donné jour à la troisième génération de tests informatiques impossibles à réaliser en tests traditionnels. Le masquage visuel rétroactif est un bon exemple de cette catégorie de tests, la perturbation d'un stimulus visuel par la survenue d'un second stimulus constitue le phénomène de masquage visuel rétroactif, expliqué par Sebban et Debouzy (1983). « Un premier stimuli est codé par un ensemble rétine et corps genouillé. Ce codage parallèle est stocké dans "un tampon" situé au niveau du cortex strié. Les processus d'identification séquentiels des résultats du codage entrent en jeu par comparaison avec des données mémorisées. Pendant ce temps d'identification, le 2e stimulus est vu et codé, les résultats de ce deuxième codage allant "écraser" ceux du premier dans la mémoire tampon. On mesure donc la rapidité du processus élémentaire du traitement de l'information visuelle. » Ce test semble être reproductible et avoir un fort pouvoir discriminant en fonction de la gravité de l'affaiblissement intellectuel.

Les inconvénients

La lourdeur logistique : chaque test nécessite un matériel spécifique et un logiciel identique et de la même version. Ces tests sont difficilement transposables d'un système à l'autre, pour des raisons de langage informatique, mais aussi de

vitesse du système, par exemple un même test sera beaucoup moins rapide et donc non valide selon que l'écran est traditionnel ou matriciel.

Si les ordinateurs portables rendent possibles une évaluation à domicile, leur coût n'est pas négligeable.

Le risque de non-fonctionnement peut toujours exister alors qu'il est nul avec les tests traditionnels : un support technique est à recommander, quand on réalise une recherche multicentrique intégrant des tests informatisés.

L'apprentissage et les progrès réalisés par les sujets sont beaucoup plus importants que dans les tests traditionnels, à tel point que beaucoup d'auteurs de tests se servent de ces tests pour une rééducation cognitive.

Il est important pour comparer différents temps de «gommer» l'apprentissage par un entraînement des sujets, pour qu'ils atteignent leur meilleur niveau de performance.

DESCRIPTION DES BATTERIES PUBLIÉES ET APPLICATIONS

Cette liste n'est pas exhaustive, les tests cités sont donnés à titre d'exemple; il existe de nombreuses batteries informatisées : chacun réalisant ses propres tests.

Systèmes experts et tests de personnalité

– **Mini-MULT :** questionnaire réduit à 71 items du MMPI en passation classique, avec une correction et une exploitation entièrement automatisées.

Application : étude de la personnalité normale et pathologique. Éditions du centre de psychologie appliquée (ECPA).

– **EXPLORE :** test de passation classique avec une correction et une exploitation entièrement automatisées.

Application : cette batterie explore 5 aptitudes intellectuelles et 15 traits de personnalité (16 PF de Cattell). Éditions : ECPA.

– **SOSIE :** il s'agit de questionnaires automatisés avec une correction et une exploitation automatisées par le Minitel.

C'est la réunion de trois inventaires de personnalité de L. V. Gordon :

– le GPP-I qui évalue 9 traits de personnalité,
– le SIV qui évalue 6 valeurs interpersonnelles,
– le SPV qui évalue 6 valeurs personnelles.

Le temps de passation est d'environ 45 minutes.

Application : Sosie est un outil qui aide le formateur à recenser les besoins de formation. Éditions ECPA.

– **Le D5D :** test de personnalité entièrement automatisé sur PC ou compatible; il s'agit d'un inventaire de 55 descripteurs à choix forcé, d'une passation inférieure à 15 minutes, avec une disponibilité immédiate des résultats avec affichage instantané des scores et des profils, réalisé par Jean-Pierre Rolland et Jean-Luc Mogenet.

Issu de la première validation française du modèle des Big Five[1], le système D5D est un ensemble modulaire informatisé qui évalue l'individu selon les 5 dimensions fondamentales de la personnalité. La rapidité de passation et de lecture fait de ce système un instrument de dialogue entre les différents partenaires d'une opération de sélection. Éditions ECPA.

Batteries neuropsychologiques classiques informatisées

– **CAMDEX** (Roth *et al.*, 1986) : *The Cambridge Mental Disorders of the Elderly Examination.* Cette batterie est un instrument permettant d'effectuer un diagnostic précoce de la maladie d'Alzheimer.

Matériel : la batterie de tests informatisés n'étant qu'un des éléments du bilan qui consiste en un interview structuré du sujet, des tests cognitifs automatisés en relation avec les critères DSM-III de démence comprenant l'orientation, le langage, la mémoire, les praxies, l'attention, la pensée abstraite, la perception et le calcul.

Application : évaluation clinique et cognitive des malades déments.

– **Batteries de tests du CDR** (Simpson *et al.*, 1991; Wesnes *et al.*, 1991). Cette batterie a pour but d'évaluer l'efficacité mentale d'un sujet dans différentes opérations. Elle a été validée et étalonnée chez des sujets adultes jeunes et âgés sains et dans les différentes pathologies, il existe une base de données des variations-type induites par différents médicaments psychotropes. Plusieurs niveaux de difficulté et formes parallèles existent. Elle existe dans la plupart des langues européennes.

Matériel : PC, avec carte couleur et matériels confidentiels.

Elle contient un temps de réaction simple, un temps de réaction au choix, un test de vigilance (attention aux nombres), une reconnaissance immédiate et à long terme de mots, d'images et de visages, une exploration de la mémoire immédiate des chiffres.

Application : la psychopharmacologie; cette évaluation est plus adaptée aux phases I et II : il s'agit d'une batterie à louer auprès du CDR qui assure l'intendance complète des tests : prêt de matériel, formation des investigateurs, exploitation des données. Mais on ne peut pas disposer librement de ces tests.

– **ECO** (évaluation cognitive par ordinateur) (Ritchie *et al.*, 1993), batterie de tests pour évaluer les fonctions intellectuelles.

La batterie comprend : un temps de réaction simple, une exploration de l'attention auditive, de l'attention visuelle, de la mémoire primaire (rappel verbal immédiat, registre visuo-spatial, tâches doubles), de la mémoire secondaire (rappel verbal différé, rappel visuel différé, un rappel de récits, mémoire implicite), une exploration de l'organisation et de l'analyse visuo-spatiale (formes géométriques, associations fonctionnelles et sémantiques, suites logiques), des performances visuo-spatiales (praxies gestuelles et constructives), du langage (évocation lexicale, dénomination, phonomorphologie, compréhension phonémique), décision lexicale (Anonyme Eco, 1993; Prud'homme *et al.*, 1990).

1. Le modèle des Big Five décrit la personnalité selon 5 dimensions bipolaires : introversion-extraversion, amabilité positive/négative, consciencieux-peu consciencieux, stabilité émotionnelle-anxiété et ouverture aux expériences. Voir également le chapitre de Michel Huteau, *L'approche de la personnalité par les traits*, dans ce même ouvrage (note des éditeurs).

C'est une batterie validée et étalonnée sur 335 personnes âgées saines, les validations dans différentes pathologies sont en cours.

Matériel : Macintosh classique, logiciel écrit en langage Hypercard 2,1 demande un écran tactile : carte Mactouch. Ce logiciel est diffusé gratuitement.

Applications : évaluation des fonctions cognitives de sujets âgés, sains lors d'enquêtes épidémiologiques par exemple, ou atteints de diverses pathologies cérébrales : troubles de la mémoire liés à l'âge ou de maladie d'Alzheimer.

– **Batterie de Fondarai *et al.*** Cette batterie est l'automatisation d'une trentaine de tests validés en version papier-crayon et couvrant un large éventail d'exploration du psychisme.

Ils sont utilisés en clinique gériatrique (Micas *et al.*, 1993), psychiatrique, traumatologie, pédiatrique et essais cliniques (Crocq *et al.*, 1985), ainsi qu'en sélection d'aptitude, aux sports de compétition (Missoum et Fondarai, 1983) à la conduite automobile mais aussi en docimologie.

Pour exemples : il s'agit de temps de réaction visuel et auditif, test des carrés (Zazzo adapté), test de reconnaissance des mots, test de Stroop, test d'empan mnésique de Wechsler, test des couleurs, masquage visuel rétroactif,... Validation sur sujets jeunes et âgés sains et malades (démences d'Alzheimer).

Matériel : Apple Macintosh ou compatible MS-DOS.

– **Batterie de Yesavage *et al.*** (1983) : il s'agit d'un test d'attention qui est un test de poursuite, un test de mémoire, le Buschke, et un test de performance motrice qui est un simulateur de vol.

Matériel : Apple I (48 Koctets), plus une carte de langage Pascal avec 16 Koctets supplémentaires pour un total de 64 Koctets, et deux lecteurs de disquettes supportant les données et les programmes, ainsi qu'un système de mesure du temps.

Application : le test d'attention est utilisable en psychopharmacologie, le test de mémoire est destiné à la recherche de la connaissance de base en psychologie et le test de performance peut servir à évaluer des performances motrices complexes.

Tests originaux informatisés « cognitivistes et/ou écologiques »

– **Batterie BCC** (bilan cognitif chronométrique) de J. Poitrenaud (Deltamed, 1995) et C. Sebban (1983). Cette batterie évalue la durée d'opérations mentales élémentaires à partir de la mesure des temps de réaction à une série de situations de difficulté croissante. Elle comporte aussi des tests d'évaluation des capacités psychomotrices.

Batterie de tests : Flicker, tests de masquage visuel rétroactif, test de Poon (1983)-Posner (1975), test de poursuite, calcul mental, évaluation de l'effet d'une tâche perturbatrice sur l'empan mnésique, test de mémoire à moyen terme, temps de réaction stimulus auditif, et visuel (temps de passation variant selon les tests choisis).

Matériel : PC compatible (386 minimum), avec table à digitaliser.

Application : la recherche clinique et pharmacoclinique (phases I et II), lorsqu'on veut avoir le profil d'une molécule sur les étapes du traitement de l'information.

– **Batterie modifiée informatisée Mac-Q**. Ce sont des tests psychométriques « testant la mémoire écologique », informatisés et mis au point par Crook *et al.* (1986) présentés dans l'ordre suivant : objets rangés à retrouver, association nom-visage : présentation de 14 visages de personnes énonçant leur prénom et la ville où elles habitent, mémoire immédiate puis différée; composer un numéro de téléphone;

reconnaissance du nouveau visage dans une cohorte de visages déjà apparus. Objets égarés à replacer, association nom-visage : à reconnaître (temps de passation : 60 min). Ces tests ont été validés chez des sujets âgés sains, présentant un trouble de la mémoire lié à l'âge et une démence d'Alzheimer.

Matériel : ces tests sont à louer à la *Memory Assessment Clinics, Inc.* qui se charge de fournir le matériel : lecteur de disque laser vidéo et ordinateur PC, et assure l'intendance.

Application : cette batterie a été réalisée pour l'évaluation des troubles de la mémoire liés à l'âge, entité nosologique décrite par les auteurs sous le nom d'AAMI (Age Associated Memory Impairment) puis d'AACI (Age Associated Cognitive Impairment).

– **Batterie de mémoire en simulation vidéo (CREAM)**. Cette batterie créée par Hibert *et al.* (1994) est un test de mémoire écologique non informatisé mais composé d'une vidéo de six séquences simulant des scènes de la vie quotidienne, comme le rangement d'objets dans des pièces, une observation de jeux et d'habits d'enfants, la reconnaissance de lieux déjà vus (magasins), un apprentissage topographique d'une ville lors d'une visite guidée en simulant le déplacement d'une voiture; un apprentissage des noms et de la qualité de dix personnages présentés dans une séance de travail filmée. Une bande s'inspirant du même type de situation est utilisée pour la rééducation cognitive.

Matériel : un magnétoscope : exploitation traditionnelle des résultats obtenus individuellement ou en groupe.

Application : l'évaluation et la stimulation de la mémoire dite écologique en individuel ou en groupe.

– **Batterie MP6-St Etienne** (Thomas-Anterion *et al.*, 1993). Batterie destinée à étudier l'apprentissage visuo-moteur pour l'étude de la mémoire procédurale. Cette batterie est composée de différents tests : tableau noir, les taches, le barrage des traits, le labyrinthe, les figures géométriques, l'écriture contrainte facile. Ces tests ont été validés chez des sujets adultes sains jeunes et âgés.

Matériel : tests créés par Thomas-Anterion sur Macintosh avec le programme Mac Paint.

Application : étude de la mémoire procédurale, par exemple dans la maladie de Parkinson, ou les traumas crâniens.

CONCLUSION

Des batteries de tests informatisés sont dès maintenant à la disposition des cliniciens et des chercheurs en neuropsychologie, quel que soit le domaine.

Ces évaluations sont plus adaptées à certaines situations, où il existe des modifications cognitives attendues à mesurer, comme l'évaluation clinique des psychotropes par exemple.

Cependant, plus encore qu'avec les tests traditionnels, il semble important avant de s'en servir d'en contrôler la validation et l'étalonnage.

BIBLIOGRAPHIE

Batterie d'évaluation de la mémoire : Psycholab - Côte-des-Neiges, 1992.

ECO : évaluation cognitive par ordinateur : évaluation neuropsychométrique de l'altération intellectuelle chez le sujet âgé. IPSEN, 1993.

CARR A. C. *et al.* — Automated Cognitive Assessment of Elderly Patients. A Comparison of Two Types of Response Device. *Br. J. of Clin. Psychol., 25*, 305-306, 1986.

CROCQ L., BUGARD P., FONDARAI J. — Les échelles d'évaluation et l'informatique, bases des études multicentrées. *Le Concours Médical, 28* (supplément), 83-91, 1981.

CROCQ L., FONDARAI J., KOHLER F. — *Tests informatiques sur micro-ordinateur.* Congrès de Psychiatr. et de Neurol. de langue française, Luxembourg, Masson, 1984.

CROCQ L., RIGAL J., CROCQ M. A. *et al.* — Effets résiduels des benzodiazépines hypnotiques sur la vigilance et l'efficience au réveil. Essai comparatif contre placebo du loprazolam, du triazolam et Flunitrazépam au moyen de tests psychométriques informatisés. *La Vie Médicale, 66*, 1443-1449, 1985.

CROOK T. H., SALAMA M., GOBERT J. — *A Computerized Tests Battery for Detecting and Assessing Memory Disorders, Senile Dementia : Early Detection.* Ed. A. BEST *et al.*, John Libbey Eurotext, 1986.

HIBERT O., LAURENT B., CROISILE B. *et al.* — Évaluation d'une batterie de mémoire en simulation vidéo de vie quotidienne. *Ann. Réadapt. Méd. Phys., 37*, 265-274, 1994.

MICAS M., FONDARAI J., VELLAS B. — Détection précoce des troubles cognitifs de la personne âgée. *La Revue de Gériat., 17*, 469-476, 1992.

MISSOUM G., FONDARAI J. — Informatique et pratique des APS. *EPS, 184*, 23-26, 1983.

MAULUCCI R. A., ECKHOUSE R. H. — The Use of Computers in the Assessment and Treatment of Cognitive Disabilities in the Elderly : a Survey. *Psychopharmacology Bulletin, 24*, 4, 557-563, 1988.

PATER V., POITRENAUD J., LE ROCH K. *et al.* — Effets différentiels de l'âge sur la fréquence critique de fusion et le masquage visuel rétroactif (à paraître, Presse Médicale).

POON L. W. — Application of Information-Processus Technology in Psychological Assessment. In CROOK T., FERRIS S., *Assessment in Geriatric Psychopharmacology.* Mark Powley Ass., New Canaan, Connecticut, 187-201, 1983.

POSNER M. J., SNYDER C. R. R. — Facilitation and Inhibition in the Processing of Signals. In RABBITT P. M., DORNIC S. *Attention and Performance*, Academic Press, London, 5, 669-682, 1975.

PRUD'HOMME M., FONDARAI J. *et al.* — Recherche de moyens d'études en psychopharmacologie clinique de l'âge. Application à l'évaluation thérapeutique. *Psychologie Médicale, 22*, 9, 871-882, 1990.

RITCHIE K., ALLARD M., HUPPERT C. *et al.* — Computerized Cognitive Examination of the Elderly (ECO) : the Development of a Neuropsychological Examination for Clinic and Population Use. *Int. J. Geriat. Psych., 8*, 700, 1993.

ROTH M., TYM E., MOUNTJOY C. Q. *et al.* — CAMDEX. A Standardized Instrument for the Diagnosis of Mental Disroder in the Elderly with Special Reference to the Early Detection of Dementia. *Br. J. Psychiatry, 149*, 698-709, 1986.

SEBBAN C., DEBOUZY C. — Le masquage visuel rétroactif. Utilisation dans l'évaluation et les essais thérapeutiques de la démence sénile. *La Presse Médicale, 12*, 48, 3150-3153, 1983.

SIMPSON P. M., SURMON D. J., WESNES K. — The Cognitive Drug Research System for Demented Patients : a Validation Study. *Intern. J. of Geriat. Psychiat., 6*, 95-102, 1991.

THOMAS-ANTERION C., LAURENT B., FOYATIER-MICHEL N. — Mémoire procédurale et ordinateur : évaluation chez des sujets normaux et dans la maladie de Parkinson. *Rev. de Neuropsych., 3*, 1, 63-85, 1993.

THOMAS-ANTERION C., FOYATIER-MICHEL N., GIRAUD S. *et al.* — Mémoire procédurale et ordinateur : évaluation chez 11 traumatisés crâniens. *Annales de Réadaptation et de Médecine Physique*, *36*, 419-427, 1993.

WENS L., BARO F., D'YDEWALLE G. — The Information Processing Approach in Clinical Memory Assessment. In *Classification and Diagnosis of Alzheimer Disease*. WHO, HOGREFE and HUBER (eds), Toronto, Lewiston, N.Y., Bern, Stuttgart, Göttingen, 103-112, 1989.

WESNES K., SIMPSON P., CHRISTMAS L. — The Assessment of Human Information-Processing Abilities in Psychopharmacology. In *Human Psychopharmacology, Measures and Methods*, vol. 1, Wiley, Chichester, 79-91, 1991.

WEST R. L., CROOK T. H. — Age Difference in Everyday Memory : Laboratory Analogues of Telephone Number Recall. *Psychology and Aging*, *5*, 520-529, 1990.

YESAVAGE J. A., TINKLENBERG J. R. — Tests d'attention, de mémoire et de performances motrices sur ordinateur appliqués à la recherche en gériatrie. *La Presse Médicale*, *12*, 48, 3170-3172, 1983.

6

LES ÉCHELLES D'ÉVALUATION DANS LES ESSAIS CLINIQUES DE PSYCHOTROPES

par A. SPRIET et T. DUPIN-SPRIET

L'évaluation clinique des effets d'un psychotrope impose une mesure de l'intensité de phénomènes ressentis ou observés, mais non directement quantifiables par un appareil de mesure par exemple. À cette fin, une échelle d'évaluation est utilisée.

Le principe d'une échelle consiste à *codifier le jugement d'un cotateur averti*, en tenant compte de ce qu'exprime le patient, de son comportement et de l'information provenant de l'entourage. Ce jugement est décomposé en donnant une note à chacune des facettes du syndrome. Il s'agit du codage de l'opinion de l'investigateur et non d'un questionnaire sur lequel on noterait les réponses du patient. À partir de ces notes, on recompose un score global qui représente la gravité du syndrome étudié.

CONSTRUCTION D'UNE ÉCHELLE

Afin de répondre à son objectif, qui est de donner une cotation fiable, crédible, et interprétable, une échelle ne doit pas être improvisée, ni « bricolée » mais construite selon une stratégie rigoureuse, puis validée pour l'utilisation envisagée.

Postulats de base (Nunnaly, 1970; Hamilton, 1987; Guelfi et Bobon, 1989)

Le mode de construction d'une échelle postule que le phénomène à mesurer est un continuum, que sa gravité est proportionnelle à celle de l'ensemble des symptômes pris en compte et au nombre de ceux qui se manifestent. Ainsi, un patient, qui présentera plus de symptômes qu'un autre (à égalité de leur intensité), sera considéré comme plus grave.

Beaucoup d'échelles postulent, en outre, implicitement que le symptôme dont un malade parle spontanément est plus grave que celui qui a été retrouvé par l'interrogatoire. Le libellé des formulaires utilisés pour ces échelles en tient compte.

Les échelles postulent également que tous les symptômes étudiés évoluent dans le même sens (ils ont tous des corrélations positives).

La description des grades de chaque symptôme doit correspondre à des gravités croissantes ou décroissantes.

Enfin, la construction d'une échelle suppose que le score global (obtenu en général simplement par addition de chacun des scores des symptômes) est proportionnel à la gravité, quelle que soit sa composition. En d'autres termes, deux patients qui ont des notes égales mais de compositions différentes sont supposés avoir la même gravité. Ce dernier postulat ne correspond vraisemblablement pas à la réalité exacte, mais il n'est pas contredit par la pratique des essais cliniques.

Étapes

Élaboration de la première version

La première étape de construction d'une échelle consiste à rassembler des descriptions cliniques de la pathologie étudiée, avec une analyse sémiologique complète. Une collection d'items (éléments distincts, compréhensibles et univoques) est réalisée. Pour chacun, on décrit plusieurs stades de gravité : descriptions ordonnées, non ambiguës, et en nombre restreint (souvent de 3 à 5). Un nombre de grades trop peu important perd une partie des nuances. Un nombre de grades trop important est source de confusion. À partir de cette collection d'items, une première sélection élimine les symptômes trop peu fréquents, et ceux qui sont redondants (exprimant le même phénomène avec des mots différents).

La rédaction doit éviter : les questions négatives (*non, ne pas, sans, absence de,...*), pour lesquelles une réponse négative peut être trompeuse et les questions doubles (*ou, et, ni*), pour lesquelles une réponse peut être ambiguë.

Les grades sont supposés représenter des degrés également espacés de gravité. Cela est impossible à prouver en pratique, et l'on peut indirectement approcher cette exigence en étudiant des adverbes d'intensité perçus subjectivement comme également espacés par une population témoin.

L'évaluation peut être absolue (intensité à l'instant même ou depuis 24 ou 48 heures, ou pendant les 7 derniers jours), ou relative (par exemple «*depuis la dernière fois*»).

Une première version de l'échelle est ainsi construite. Elle est utilisée sur un échantillon de sujets présentant la pathologie étudiée avec des gravités différentes. Sur ce premier corpus de données, une analyse repère les items fortement ou faiblement corrélés avec le score total. Une double cotation : «test» puis «retest» évaluant les malades deux fois permet d'apprécier les items reproductibles de l'échelle.

Les résultats de ces analyses permettent de modifier progressivement la première version de l'échelle par retraits et ajouts d'items et remodelage de la rédaction. Le même processus est répété sur des patients différents jusqu'à ce qu'une version simple et efficace soit obtenue : l'échelle définitive.

Les items et leurs grades doivent être monotones (la gravité cotée doit correspondre à de réelles intensités croissantes), ne pas présenter d'effet plafond (les notes les plus fortes correspondent en moyenne à des patients plus graves que les notes immédiatement inférieures).

Échelle définitive et sa validation

L'échelle définitive (ou supposée telle) est alors soumise à d'autres analyses, pour tester ses propriétés métrologiques :

– cohérence interne (corrélation item-total : tous les items évoluent en moyenne dans le même sens pour des gravités croissantes de la pathologie traitée);

– fidélité inter-juges (le même malade coté par deux évaluateurs obtient la même note);

– validité (l'échelle mesure bien le symptôme qu'elle est censée évaluer). Plusieurs critères de validité peuvent être utilisés :

– validation empirique en comparant les résultats de l'échelle à une classification clinique indépendante des sujets de gravité croissante;

– validation concourante en comparant les résultats d'une nouvelle échelle à ceux d'un instrument plus classique;

– validation prédictive (si l'échelle est destinée à prévoir le pronostic du patient) en la comparant à l'évolution réelle, connue ultérieurement.

La structure de l'échelle est étudiée par des analyses factorielles, en particulier l'analyse en composantes principales, qui dégage d'un ensemble de variables multiples (les différents items), un certain nombre de facteurs, moins nombreux, représentant le mieux possible l'information sur les différences observées entre les sujets (variance). Ces facteurs doivent ensuite être interprétés cliniquement (par exemple les facteurs d'une échelle de dépression pourraient être l'humeur, les troubles somatiques, l'anxiété, le sommeil et le ralentissement).

Modification d'une échelle

Il ne faut pas, sans précautions, supprimer, ajouter ou modifier des items, même si cela paraît évident dans un contexte donné. En effet, la validité et la structure de l'échelle peuvent en être affectées de façon inconnue, et les résultats d'essais cliniques l'utilisant pourront être contestés.

De la même façon, on ne peut utiliser une sous-échelle (sous-ensemble d'items) que s'il s'agit d'un groupe autonome dont on a prouvé indépendamment la validité pour mesurer un facteur cliniquement identifié.

La traduction d'une échelle pose des problèmes redoutables, car les contresens sont faciles même pour des traducteurs chevronnés. En effet, il faut tenir compte non seulement de la langue, mais de la culture locale, voire de l'école psychiatrique, pour comprendre exactement et ré-exprimer un concept en psychiatrie. Idéalement, une nouvelle validation devrait être faite sur un échantillon représentatif de la population parlant la langue de la version traduite. Au minimum, une re-traduction (traduction de la première traduction pour retourner vers la langue originale) permet de s'apercevoir des contresens graves.

UTILISATION DANS LES ESSAIS CLINIQUES (Gram, 1991; Spriet *et al.*, 1993)

Choix d'une échelle

Lorsque l'on en a plusieurs à sa disposition pour étudier un syndrome donné, le choix prendra en considération :

– L'objectif : l'échelle sera-t-elle utilisée comme critère principal (pour lequel il faut être le plus exigeant possible sur la validation) ou accessoire? Sera-t-elle utilisée pour la sélection de patients éventuellement très perturbés, pour lesquels l'expression spontanée et l'interrogatoire seront de peu de secours?

– Cherche-t-on une cotation d'un symptôme très général, ou très spécifique : gravité d'une schizophrénie *versus* signes positifs ou signes négatifs (Andreasen et Grove, 1986); symptomatologie générale de la dépression *versus* ralentissement dépressif?

– Le protocole : la durée du traitement et la fréquence des examens conduiront-elles à répéter souvent l'évaluation?

– Le personnel disponible (Raskin et Crook, 1976) : est-il qualifié pour remplir une échelle élaborée? Sinon, pourra-t-il suivre un entraînement adéquat?

– Les destinataires de l'essai clinique : les autorités d'enregistrement aux États-Unis préfèrent certaines échelles, ceux des autorités européennes en préfèrent d'autres (voire même plusieurs pays européens ont des choix particuliers...).

– Choix disponible : connaît-on toutes les échelles disponibles (bibliographie, catalogue et recueils d'échelles, traductions disponibles, validation de ces traductions) (Cottraux *et al.*, 1985; Guelfi, 1993; Van Riezen et Segal, 1988).

– Existe-t-il un instrument facile à utiliser, à comprendre, et ne prenant pas trop de temps à l'évaluateur?

Cotation

L'échelle doit être remplie par un cotateur averti, après un entretien avec le patient. Plusieurs modalités existant, le choix doit être précisé dans le protocole afin qu'il soit constant et ainsi que les données recueillies sur différents patients et à différents moments soient comparables entre elles :

– *entretien libre* (le psychiatre laissant le patient s'exprimer spontanément);

– *entretien structuré* (permettant de systématiser l'exploration des symptômes et l'ordre dans lequel ils sont explorés), sans toutefois poser des questions trop directives induisant les réponses;

– *entretien semi-structuré* : laissant le patient s'exprimer et l'orientant vers les questions sur lesquelles il ne s'est pas exprimé.

En outre, on a parfois recours à la méthode « temps aveugle » (Mormont *et al.*, 1984), dans laquelle les malades, ayant donné leur consentement pour cette opération, sont enregistrés sur magnétoscope. Leurs entretiens successifs sont visionnés dans le désordre par le cotateur, à qui on a masqué la chronologie des examens. Les passages dans lesquels le patient fait allusion à cette chronologie sont brouillés.

Cotateur

Le cotateur averti doit savoir écouter le malade, interpréter ses dires et ses comportements, recouper des informations d'origine différente. Il doit bien connaître la pathologie étudiée, et savoir reconnaître les symptômes légers, trompeurs, ou indirects.

Le cotateur doit connaître l'instrument et doit être conscient des tendances de réponses qu'il pourrait avoir et qui pourraient fausser les résultats : tendance à l'atténuation, erreur de proximité (donnant des réponses voisines à des questions voisines), tendance extrême (caricaturant les réponses en tout ou rien), ou centrale (donnant à tous les items des réponses voisines du grade médian), et enfin dérive (tendance à donner au patient vu régulièrement des scores qui s'améliorent dans le temps).

L'entraînement des cotateurs permet de leur faire prendre conscience de ces tendances et ainsi de les réduire. Il se fait le plus souvent en séances de groupe, au cours desquelles sont projetés des enregistrements de patients (vrais ou simulés). Une séance de cotation efficace comprend : un protocole écrit de cotation, une sélection des participants dont au moins un est très expérimenté dans la cotation. Les cotateurs doivent être installés de façon à voir, entendre et écrire correctement. La projection doit être de bonne qualité technique (en particulier le son). Les participants reçoivent des consignes de silence et d'indépendance. Ils remplissent leur échelle complètement avant la discussion aboutissant au consensus. Les cotateurs « extrêmes » sont identifiés. Le plus souvent, ils prennent conscience de leurs tendances (par exemple réponses exagérées qu'ils peuvent modifier au cours d'exercices ultérieurs) ou d'une interprétation non consensuelle d'un item qu'il faudra clarifier.

Chronologie des mesures

Une échelle répétée dans le temps permet d'étudier l'évolution d'un patient donné. La chronologie des mesures doit être la même pour tous les patients (à une petite marge de tolérance près qui est précisée dans le protocole), et les temps doivent être choisis de façon à déceler une différence de cinétique d'effet (un traitement agissant plus rapidement ou de façon plus prolongée qu'un autre).

Formulaire de recueil

Le formulaire de recueil d'une échelle n'est pas un questionnaire (dans lequel on enregistrerait sans interprétation les réponses des patients), mais un document destiné à recevoir l'opinion de l'évaluateur à l'issue d'un entretien avec le patient. Pour une même échelle, il en existe souvent plusieurs versions :

– formulaire résumé pour cotateurs avertis,

– formulaire détaillé comportant la description de chaque item pour les cotateurs moins familiarisés avec l'instrument.

En outre, le formulaire doit rappeler les conditions de l'examen (sur quelle durée des symptômes doivent-ils être évalués, quel type d'entretien par quel type de cotateur), et éventuellement la référence de la version (si elle a été modifiée, par exemple en ce qui concerne le nombre d'items, le codage des grades, ou les traductions).

Contrôle des réponses

Pour être utilisable, une échelle doit être correctement remplie, il importe donc que le clinicien ou un assistant de recherche clinique vérifie qu'elle est complète, que les réponses sont lisibles et sans ambiguïtés, qu'il existe une seule réponse par item, que les totaux et sous-totaux sont justes. Plus difficile est la recherche de la cohérence, soit à l'intérieur d'une échelle (il pourrait y avoir des réponses apparemment contradictoires nécessitant confirmation entre deux items de signification voisine ou fortement corrélés), ou cohérence inter-échelles (les items exprimés de façon voisine en deux échelles remplies pour le même patient donnant des réponses nettement contradictoires nécessitent confirmation). La détection de réponses méritant confirmation permet de détecter des confusions de patients, de documents, ou même un malentendu de compréhension de l'échelle.

Analyse des résultats

L'analyse statistique des résultats des scores d'une échelle fait appel aux méthodes classiques d'analyse des résultats d'un essai clinique en tenant compte des facteurs suivants :

– si le score peut être considéré comme un continuum, un test paramétrique est habituellement utilisé ;

– si la cotation est manifestement discontinue, un test pour catégories ou pour classification ordonnée peut être utilisé.

Un score est habituellement calculé en additionnant toutes les notes, mais l'analyse d'items séparés permet d'étudier les profils des patients et des réponses au traitement, en général à titre exploratoire.

Des méthodes d'analyse multivariée tenant compte des items et de leurs covariances sont parfois utilisées, de même que la méthode consistant à calculer sur les échelles initiales (avant traitement) des facteurs par analyse en composante principale, et à recalculer sur les scores après traitement ces mêmes facteurs en utilisant les poids de l'analyse initiale.

Des sous-ensembles d'items (sous-échelles) sont aussi parfois utilisés à titre exploratoire.

Le score initial peut être utilisé de deux façons : soit en le soustrayant de tous les scores ultérieurs (ainsi l'on suppose que le nombre de points gagnés présente le même intérêt quel que soit le niveau de départ), ou utilisé comme dénominateur d'une fraction, permettant de calculer le pourcentage d'amélioration (si l'on suppose qu'un gain donné est plus important si le score initial est plus bas).

AUTO-ÉVALUATION
(Bobon et Bobon-Schrod, 1974; Luria, 1975; Leserman et Koch, 1993)

Une échelle d'auto-évaluation est un instrument comportant d'une part des consignes, instructions données par écrit et expliquées oralement aux patients pour leur expliquer comment coter eux-mêmes leurs symptômes, et d'autre part un questionnaire permettant de recueillir leurs réponses de façon standardisée.

Consignes

Elles doivent être courtes, compréhensibles (surtout pour des patients peu familiarisés avec la cotation et l'introspection), précises avec, si possible, des exemples explicatifs, et claires sur la chronologie du symptôme exploré. Il faut également préciser au patient qu'une et une seule réponse doit être fournie pour chaque question.

Le formulaire

Il peut être de plusieurs types :

– *Questions à choix multiples* : les questions doivent être formulées de façon simple et claire, la signification de toutes les réponses proposées ne doit comporter aucune ambiguïté, les grades doivent être perçus par le patient comme distincts, et la présentation de tous les items doit être homogène pour en faciliter la compréhension et le remplissage par le patient et la vérification par l'investigateur.

– *Échelle visuelle analogique* (Luria, 1975) : dans cette présentation, on demande au patient de se situer sur un continuum représenté par une ligne tracée sur une feuille (en général horizontale, parfois verticale) dont les deux extrémités sont caractérisées par des expressions extrêmes et sémantiquement opposées, relatives au symptôme que l'on tente d'évaluer. Exemple : «*je n'ai jamais eu aussi faim*» et «*je n'ai pas faim du tout*». Le sujet se situe sur ce continuum par un petit trait. Il n'est pas recommandé de faire figurer des repères intermédiaires sur une échelle à remplir, sauf un repère central dans le cas où l'item serait relatif (*mieux – comme avant – moins bien*).

Il existe une polémique concernant la possibilité qu'un patient remplisse une échelle au vu de ses réponses précédentes. En fait, en voyant ou en ne voyant pas la réponse précédente, on ne répond pas à la même question. Si le patient voit sa réponse précédente, il a tendance à coter son évolution. S'il ne la voit pas, il évalue son état. Il faut en être conscient et le systématiser pour tous les patients d'un même essai.

– *Liste d'adjectifs* (Bobon et Bobon-Schrod, 1974) : dans ce modèle, des couples d'adjectifs sémantiquement opposés, souvent sans relation directe avec la description de la pathologie, sont présentés au patient, en lui demandant de faire un choix.

Conditions d'utilisation

Comme pour toute méthode d'auto-évaluation (cela est valable pour l'auto-mesure de la pression artérielle ou le dosage de la glycémie par le patient lui-même), les échelles d'auto-évaluation ne sont valables que si trois impératifs sont respectés :

– *Compréhension* : il ne faut pas inclure dans un essai avec auto-évaluation des malades trop perturbés, des malades ayant des difficultés pour lire, écrire ou comprendre la langue dans laquelle est rédigé le formulaire. Quoi qu'il en soit, les consignes et les échelles doivent être rédigées dans un langage simple, et les mots savants, les tournures précieuses, les phrases longues, le jargon médical ou les expressions, dont le sens médical est différent de celui du public (manie, délire, démence), sont à proscrire.

– *Adhésion* : il ne faut pas inclure dans une auto-évaluation des malades ne pouvant pas analyser correctement leur symptomatologie, ceux qui sont peu motivés,

ceux qui ont tendance à tourner en ridicule le procédé, ceux qui sont révoltés par la formulation de certaines questions, ou qui pourraient être gênés par certaines d'entre elles jugées absurdes ou déplacées dans le contexte de leur pathologie (évaluation de l'appétence au travail chez un chômeur ou chez un mourant). En outre, la longueur ou la répétition du formulaire à remplir peut exaspérer certains patients et compromettre la crédibilité de leurs réponses.

– *Contrôle* : il vaut mieux que le malade soit sous observation (mais non sous influence) lorsqu'il remplit son échelle d'auto-évaluation.

Dès qu'il l'a remplie, ou à défaut, lors de la prochaine visite pour un patient en ambulatoire, il faut vérifier si l'échelle est complète (y a-t-il des oublis, ou une impossibilité de se prononcer pour certains items), cohérente (certaines réponses sont-elles contradictoires apparemment avec d'autres), n'y a-t-il pas eu alors malentendu ou confusion?

Les notations sont-elles lisibles? N'y a-t-il pas de réponses doubles, ou de malentendus pour des questions difficiles? Il ne faut pas corriger l'auto-évaluation s'il n'y a pas de malentendu ni d'ambiguïté ni même en cas de non-confirmation par l'entourage (qui considère qu'il y a eu exagération ou atténuation manifeste).

CONCLUSION

Les échelles d'évaluation sont des instruments irremplaçables pour l'étude des psychotropes. Leur difficulté d'utilisation et les situations dans lesquelles elles donnent des résultats crédibles doivent être connues. L'analyse la plus sophistiquée ne pourra jamais corriger une utilisation approximative, superficielle, ou la méconnaissance de leurs conditions d'utilisation.

Bibliographie

Andreasen N. C., Grove W. M. — Evaluation of Positive and Negative Symptoms in Schizophrenia. *Psychiatr. Psychobiol.*, *1*, 106-121, 1986.

Bobon D. P., Bobon-Schrod A. — Les listes d'adjectifs BS/BS' de Zerssen et la ligne de Scott : deux auto-évaluations de l'humeur. *Feuillets Psychiatriques de Liège*, *7/4*, 492-95, 1974.

Cottraux J., Bouvard M., Legeron P. — *Méthodes et échelles d'évaluation des comportements*. EAP, Issy-Les-Moulineaux, 1985.

Gram L. F. — Rating Scales in Psychiatry : Uses in the Assessment of Drug Effects. In : Nimmo W. S., Tucker G. T. *Clinical Measurement in Drug Evaluation*, 45-56. Wolfe, Aylesbury (UK), 1991.

Guelfi J. D., Bobon D. — Échelles d'évaluation en psychiatrie. *Encycl. Méd. Chir. (Psychiatrie)* 37200 A 10, 1-10, 1989.

Guelfi J. D. (coordinateur) — *L'évaluation clinique standardisée en psychiatrie*. Tome 1. Pierre Fabre, Boulogne-Billancourt, 1993.

Guyatt G. H. — Measuring Quality of Life : a Review of Means of Measurement in Clinical Trials of Medicines. *Pharmaceut. Med.*, *2*, 49-60, 1987.

Hamilton M. — Assessment of Psychopathology. In : Hindmarch I., Stonier P. D. *Human Psychopharmacology – Measures and Methods*, 1-17. John Wiley and sons, Chichester (UK), 1987.

LESERMAN J., KOCH G. — Review of Self-Report Depression and Anxiety Measures. *Drug Information Journal*, *27*, 537-548, 1993.

LURIA R.E. — The Validity and Reliability of the Visual Analogue Mood Scale. *J. Psychiat. Res.*, *12*, 51-7, 1975.

MORMONT C. V., FRENCKELL R. (von), LOTTIN T., MORMONT I., TROIS FONTAINES B., BOBON D. — Influence de variables temporelles (*time-blind*, *time gap*) sur l'évaluation quantitative de la psychologie. *L'Encéphale*, *10*, 3-7, 1984.

NUNNALY J. C. Jr. — *Introduction to Psychological Measurement*. McGraw-Hill, New-York, 1970.

RASKIN A., CROOK T. H. — Sensitivity of Rating Scales Completed by Psychiatrists, Nurses and Patients to Antidepressant Drug Effects. *J. Psychiat. Res.*, *13*, 31-41, 1976.

SPRIET A., FERMANIAN J., SIMON P. — L'utilisation des échelles d'évaluation en psychopharmacologie. *L'Encéphale*, *4*, 119-129, 1978.

SPRIET A., DUPIN-SPRIET T., SIMON P. — *Méthodologie des essais cliniques des médicaments*. Karger, Bâle, 1993.

VAN RIEZEN H., SEGAL M. — *Comparative Evaluation of Rating Scales for Clinical Psychopharmacology*. Elsevier, Amsterdam, 1988.

7

LA PERSONNALITÉ ET LA DISTINCTION TRAIT/ÉTAT REVISITÉES PAR LA NEUROBIOLOGIE

par R. JOUVENT

Une clinique en évolution

Tant de la part des cliniciens que de celle des chercheurs, l'évolution de la clinique est souvent négligée. Tout se passe comme si les progrès de la connaissance et de la thérapeutique ne pouvaient provenir que d'une meilleure exploration d'une réalité immuable qui trouverait son intelligibilité la plus accomplie dans le repérage des entités nosographiques et de leurs constituants.

Cette évolution de la clinique est pourtant indiscutable dans de nombreux exemples de la science médicale. En dehors de l'accès aigu, les tableaux classiques de la Goutte ont peu à peu disparu au profit de formes plus discrètes, traitées plus tôt ou focalisées (les néphropathies goutteuses ou les complications métaboliques). On est passé d'une clinique de la diathèse à une clinique des syndromes... Dans le diabète, les complications tardives, qui faisaient la une de toutes les questions d'Internat, sont peu à peu reléguées au profit de leur traitement préventif à travers les progrès de l'équilibration du diabète lui-même.

Ces changements cliniques concernent tout aussi bien la psychiatrie, les manifestations psychopathologiques subissant une évolution comparable (Jouvent et Partiot, 1994). Le tableau classique de dépression mélancolique tend à se raréfier au profit de formes débutantes dépistées précocément ou « abatardies » par des psychotropes (comme le disent les médecins des angines soignées partiellement par un traitement antibiotique). Les états maniaques sont plus rares depuis la thérapeutique par le lithium... De nouveaux tableaux cliniques apparaissent, où les facteurs iatrogènes sont prépondérants.

Un autre changement concerne la prescription des psychotropes aussi bien dans sa qualité que sa durée. Là où certaines substances étaient censées être réservées au traitement des états aigus (par exemple les antidépresseurs), les mêmes sont maintenant préconisées pour la prévention des rechutes, voire comme un traitement chronique.

Un autre phénomène tient au fait que dans la littérature internationale on assiste à un épuisement des travaux en psychobiologie et en psychopharmacologie clinique concernant les épisodes aigus. Ce désintérêt est relayé par deux tendances ; d'une part, un besoin de définir de nouvelles entités, de nouvelles sous-catégories (par exemple pour la dépression, les dépressions récurrentes brèves, les dépressions mineures...) et, d'autre part, un regain d'intérêt pour l'axe II du DSM-IV et la personnalité en général.

Des notions telles que celles de vulnérabilité, de réactivité aux substances pharmacologiques, de traitements préventifs sont l'objet de travaux croissants. Le cas du traitement «chronique» par la fluoxétine[1] y est une illustration des nouveaux problèmes posés par l'administration de substances au long cours. Certains auteurs n'hésitent pas à proposer l'idée que certaines dimensions de la personnalité seraient spécifiquement accessibles à une correction pharmacologique.

Le retour de la distinction trait/état comme abord de la question de la diachronie

Pour ces différentes raisons, le paradigme de la distinction trait-état redevient très actuel; à l'occasion de l'impact croissant des théories d'Akiskal (1990) et Cloninger (1987). En effet, et au-delà d'un regain d'intérêt pour l'axe II qu'on peut principalement attribuer à l'influence des travaux de ces deux auteurs, ces théories posent fondamentalement la question de la diachronie de l'observation clinique. Aux confins des approches structurales, différentielles et dynamiques, les perspectives du tempérament obligent en effet à définir très précisément dans la pratique expérimentale de la recherche, laquelle ou lesquelles de ces approches sous-tendent le paradigme employé.

Les psychiatres et les psychopathologues sont souvent portés à poser des alternatives dualistes. De même qu'une bonne partie des énergies s'est employée à débattre de la primauté d'une théorie globale explicative, soit le «tout mental sans cerveau», soit le «tout cerveau sans mental» (Lipowski, 1989; Szasz, 1985 in *Le Moal & Jouvent*, 1991), de même le paradigme de la distinction trait/état continue d'alimenter nombre de publications lors de l'interprétation de résultats expérimentaux.

Mais, paradoxalement, cette distinction a jusqu'alors retenu l'intérêt des psychologues beaucoup plus que celui des psychiatres. Il est logique que les psychologues, en particulier ceux dont la personnalité est le domaine, aient pu chercher dans l'opposition trait/état un modèle d'opposition entre ce que serait le «terrain» par rapport à ce qui est de l'ordre d'un épisode pathologique; mais on peut s'étonner que les psychopathologues et les psychiatres, les praticiens cliniciens n'aient pas pu remarquer dans cette notion ce qui concernait aussi bien leur pratique que leurs recherches.

LE PRÉALABLE THÉORIQUE DANS LA SAISIE CLINIQUE

Périodiquement, la psychopathologie oscille entre deux types de théories concernant la saisie et la formalisation sémiologique. L'une, d'inspiration structuraliste et médicale privilégie les différences, les discontinuités (entre le normal et le pathologique par exemple) pour mieux les formaliser; l'autre, au contraire, délaisse l'approche différentielle pour s'intéresser aux phénomènes pathologiques. Les deux approches ne sont pas, bien sûr, antagonistes mais complémentaires. Il est bien manifeste que les modèles implicites du clinicien jouent un rôle déterminant dans son évaluation : il élabore une théorie du sujet qui est à l'origine d'une modélisation progressive de l'entretien, modélisation elle-même déterminante quant au style des réponses obtenues.

---

1. KRAMER P. B. — *Listening to Prozac®*, Viking, 1993, Trad. Fr. : *Prozac®, le bonheur sur ordonnance*, First, Paris, 1994.

Il n' y a pas une vérité clinique que le clinicien devrait approcher avec la plus grande fidélité possible, mais une grande variété de sens, répétons-le. Cette polysémie de la clinique évolue avec une grande plasticité, au gré des modes et des obédiences théoriques.

Nous voudrions ici montrer comment le préalable théorique peut influencer la nature même du matériel clinique saisi lors de l'entretien ou de la passation d'un questionnaire.

C'est le cas en particulier des *inférences inductives* pouvant influencer le profil des réponses. Demander à un patient s'il éprouve un événement agréable avec plaisir entraîne immanquablement une binarisation des réponses qui se ramènent à oui ou non.

Dans l'interprétation des résultats statistiques, ce problème de la forme liée aux attitudes de réponse a déjà été relevé par plusieurs auteurs à propos de structures factorielles dans lesquelles la forme de l'item (+ ou –) jouait un rôle aussi important que celui du contenu psychologique de ces items (Pichot *et al.*, 1984). Ainsi, dans l'exemple du *State Trait Anxiety Inventory*, les différences (et donc la variance, sur laquelle se calculent les facteurs) «entre les items positifs et négatifs sont à peu près aussi grandes que les différences entre les items d'État et de Trait...» (Bernstein et Eveland, cités par Pichot *et al.*, 1984).

Revenons à l'axe II et à l'approche de la personnalité.

L'intérêt explicatif du concept de personnalité a été amplement démontré depuis la fin du XIXe siècle. Plutôt qu'à l'influence de Ribot, c'est à celles, contemporaines, de Freud et de Kraepelin qu'il faut ici se référer. Sans forcer le trait, on peut dire que la psychiatrie d'aujourd'hui demeure encore profondément marquée par cette double influence. De manière évidemment très différente, ces deux cliniciens ont établi des lignes de continuité entre personnalités normales et pathologiques, et, grâce à ce principe de continuité, ils ont su établir des rapports de causalité entre la personnalité et certains états morbides. Sans jamais se rapprocher, les deux courants de pensée se sont développés jusqu'à présent pour donner au cadre général des personnalités pathologiques l'air composite qu'on lui connaît, et dont témoigne encore leur classification sur l'axe II du DSM-II-R.

LE DÉCOUPAGE CLASSIFICATOIRE DE LA PERSONNALITÉ COMME ÉCHAPPATOIRE AU PROBLÈME DIACHRONIQUE

On notera néanmoins que les analyses différentielles de la personnalité n'ont guère apporté les éclaircissements de nature psychopathologique que l'on pouvait en attendre. Les travaux de Eysenck pouvaient donner quelques espoirs. Dans un travail maintenant ancien (1957), il proposait que l'analyse dimensionnelle qu'il avait construite puisse permettre d'évaluer l'action des médicaments psychotropes. Sans discuter en détail les résultats de cette étude, on peut s'interroger sur le principe de la validité d'une relation entre des dimensions dégagées de l'étude de la personnalité normale et l'action attendue de molécules agissant sur des systèmes neuronaux élémentaires.

Huteau (1985) propose de substituer à des analyses qui visent à décrire, à partir d'inférences statistiques inductives, les grandes lignes organisatrices de la personnalité, des processus moins complexes qui pourraient à la fois expliquer une part des

fonctions intervenant dans l'organisation de la personnalité et susceptibles d'expliquer certains processus psychopathologiques, telles l'hypothèse du locus de contrôle ou celle de la variation dépendance-interdépendance du champ. Il n'est pas exclu qu'une telle orientation de recherche puisse conduire à des études fructueuses.

Quels que soient les facteurs qui contribuent à la réalisation d'une conduite ou d'un système de conduites pathologiques, l'objet premier de la psychopathologie n'est pas d'expliquer le rôle de ces facteurs ni de choisir entre déterminants organiques, psychologiques et sociaux. Comme l'a souligné Canguilhem, il y aurait avantage à substituer au concept d'anormal celui d'anomalie. Un système dit pathologique est d'abord un ensemble de conduites qui se détache du fond du comportement. C'est là ce qui le rend visible à l'observateur comme un objet d'investigation, objet dont il s'agit d'expliquer la présence et de voir les effets sur l'ensemble de la personnalité. Ce caractère d'anomalie peut être saisi par le sujet lui-même.

C'est néanmoins dans une idéologie différente que les troubles de la personnalité nous reviennent en référence, avec une idéologie nosologiste implicite.

LE PARADIGME DU TEMPÉRAMENT

Nous ne développerons pas en détail les différentes théories du tempérament, ni la façon dont cette notion est réintroduite dans la conceptualisation des troubles psychiatriques. Un exemple très riche en est fourni par H. Akiskal (1990) à propos des troubles thymiques. Nous voulons seulement souligner ici comment une telle notion offre une alternative au problème du choix, *a priori* obligé, entre approche différentielle et perspective du changement.

Issues d'ailleurs d'un besoin classificatoire en vue d'établir une typologie des conduites, les conceptions du tempérament n'ont pu renoncer à leurs positions « catégorialistes » qu'à partir du moment où Wundt proposa une alternative dimensionnelle : l'appartenance typologique d'un individu provenait selon lui de la combinaison fonctionnelle de dimensions telles que la force de la réactivité affective et la vitesse de changement des sentiments. Il est intéressant de remarquer que c'est sous l'influence d'auteurs appartenant au champ de la physiologie, à celui de l'étude des mécanismes, comme Wundt et Pavlov, que la notion de tempérament a basculé vers une acception dynamique, reléguant la personnalité et les traits sous la coupe des statisticiens différentialistes comme Cattell ou Eysenck.

À l'heure actuelle, et toujours par opposition à une vision implicitement structurale de la personnalité, la notion de tempérament semble retrouver un intérêt nouveau. Comme l'explique M. Reuchlin, il n'y a pas de définition heuristique du tempérament. Mais globalement, l'idée générale résidait dans l'étude des différences inter-individuelles quant à la forme des conduites ou le style des réponses émotionnelles ou sociales : « Les individus se différencient quant au niveau du stimulus le plus faible susceptible de susciter une réponse [...] paraissent différer quant à un niveau optimal de stimulation, soit par le fait de systèmes modulateurs éventuellement périphériques, soit plus vraisemblablement par le fait de structures centrales plus faciles ou plus difficiles à mettre *en éveil*. Certains sujets vont donc rechercher les stimulations et d'autres les éviter, ce qui entraînera des conduites que l'on pourra qualifier d'extraverties ou d'introverties (notamment dans un contexte social)... » (Reuchlin, 1990).

Curieusement, malgré cette connotation différentielle, la notion de tempérament semble retrouver un regain auprès des physiologistes et des psychologues du changement, à propos de deux caractéristiques : l'idée d'une interaction dynamique sujet-milieu, et la compatibilité avec une approche développementale.

L'interaction entre certaines caractéristiques du sujet avec son milieu est à la base des théories actuelles du développement. On ne travaille plus sur des conduites spécifiques mais on recherche des patterns biologiques, en amont de la diversité phénotypique, qui seraient communs à des destins comportementaux très variés. Ainsi, Strelau (1991) a montré que les individus à faible réactivité ont tendance à préférer les situations à forte valeur de stimulation, et inversement.

De la sorte, un toxicomane et un deltaplaniste peuvent avoir un même tempérament (biologique, par exemple un besoin de sensations fortes), mais en fonction de facteurs environnementaux chacun a pu acquérir une spécificité de comportement.

La connotation développementale est la seconde caractéristique spécifique du tempérament par rapport à la notion de personnalité. L'idée principale réside dans l'importance des capacités parentales à répondre aux caractéristiques biologiques de l'enfant. Ceci est très visible dans l'exemple du phénomène d'ajustement réciproque : les adultes essaient d'adapter leur comportement à celui de l'enfant qui provient au début de son seul tempérament. Progressivement, le trait de tempérament s'ajuste et se développe en relation avec les réactions parentales. Ce trait ne deviendrait un facteur de comportement pathologique que si la «*goodness of fit*» n'est pas atteinte, en d'autres termes si ce trait se combine avec un mauvais ajustement entre les capacités d'adaptation et les conditions de l'environnement (Thomas et Chess, 1977).

C'est dans cette ligne qu'on peut intégrer les intéressantes constatations cliniques d'Akiskal à propos des enfants de bipolaires : le même tempérament hyperthymique sera valorisé et renforcé par les parents chez un garçon alors qu'il sera réprimé (et donc source de honte ou de culpabilité) chez une petite fille (communication personnelle).

Par extension, on pourra rattacher ces caractéristiques individuelles biologiques à des facteurs de vulnérabilité chez l'animal (Piazza *et al.*, 1989) comme chez l'homme (Menza *et al.*, 1993).

L'ÉTAT DÉPENDANT DU TRAIT

Dans la continuité de cette perspective dynamique, et contrairement à l'opposition dualiste trait/état, il nous paraît plus fructueux de privilégier les modèles selon lesquels l'état est dépendant du trait.

De cette manière, les modèles d'étude des traits devraient avoir une conséquence implicite sur la façon d'aborder l'étude des états aigus; la modélisation expérimentale est en effet très différente selon le type de théorisation de la relation hypothétique entre trait et état.

1. L'analogie positive ou identité consiste en une similitude homothétique entre le trait et la sémiologie de l'état aigu. Par exemple, dans une telle hypothèse, un sujet ayant un trait anhédonique développe un épisode affectif où dominent l'anhédonie et l'émoussement affectif. Ce type de modèle implique que les paradigmes expérimentaux utilisés pour l'étude du trait et pour celle de l'état intègrent cette idée d'homothétie.

2. La potentialisation est proche de la situation précédente mais, dans ce cas, on assiste à une majoration sémiologique des symptômes trait lors de l'état aigu.

3. L'indépendance au contraire suppose qu'il n'y ait aucune relation entre la personnalité et la nature de l'état aigu. Ce prototype de situation laisse beaucoup de liberté dans la définition des protocoles expérimentaux, mais il est d'une valeur heuristique réduite.

4. Enfin, l'interaction est le modèle qui nous semble le plus intéressant de par sa richesse autant que par la possibilité d'applications biologiques et bio-pharmacologiques.

Dans ce modèle, la nature (les mécanismes qui sous-tendent) de l'épisode aigu peut s'expliquer en partie par des caractéristiques préalables de personnalité, que celles-ci soient d'ordre sémiologique, biologique ou bio-pharmacologique. Par exemple, un sujet, ayant un déficit d'activation pouvant expliquer certains traits de tempérament comme la recherche de stimulations compensatrices, peut, s'il est privé brutalement de ces stimulations, développer un état dépressif dominé par une sémiologie déficitaire avec anhédonie et émoussement affectif (comme dans un équivalent de syndrome de sevrage).

Dans la même ligne, un autre exemple concerne les sujets ayant le même trait de défaut d'activation qui s'automédiquent par le tabagisme; la stimulation chronique par la nicotine peut, chez ces sujets, être considérée comme un traitement psychostimulant dont l'arrêt lors du sevrage tabagique peut conduire à une dépression d'allure déficitaire. Les sujets les plus susceptibles «d'utiliser» les propriétés activatrices de la nicotine pourraient être les plus susceptibles de développer un trouble dépressif de type émoussement affectif soit lors du sevrage tabagique, soit lors d'un épisode dépressif d'une autre origine.

De la même manière, les sujets les plus susceptibles d'avoir besoin des effets sédatifs de la nicotine pourraient présenter des symptômes de type impulsivité ou anxiété.

L'APPROCHE BIOPHARMACOLOGIQUE DES TRAITS

L'idée générale de ce type de recherche consiste dans l'identification de traits phénotypiques qui ne soient plus définis par des indices cliniques ou biologiques mais par la capacité à produire une réaction biologique particulière après une stimulation pharmacologique.

On sait par exemple que tous les individus ne réagissent pas de la même manière à des psychostimulants comme la nicotine ou l'amphétamine; il s'agit ici de faire l'hypothèse que cette variable inter-individuelle est sous-tendue par une réactivité biologique particulière dont on pourra vérifier le mécanisme et tenter ensuite de décrire le ou les processus enzymatiques mis en jeu par une substance pharmacologique. Nous en donnons un exemple dans la figure 1 à propos d'un trait nicotino-dopaminergique.

À moyen terme, une telle recherche devrait permettre de renouveler l'approche génétique non seulement des addictions mais de bon nombre de syndromes psychopathologiques, et ouvrir des perspectives en pharmaco-génétique.

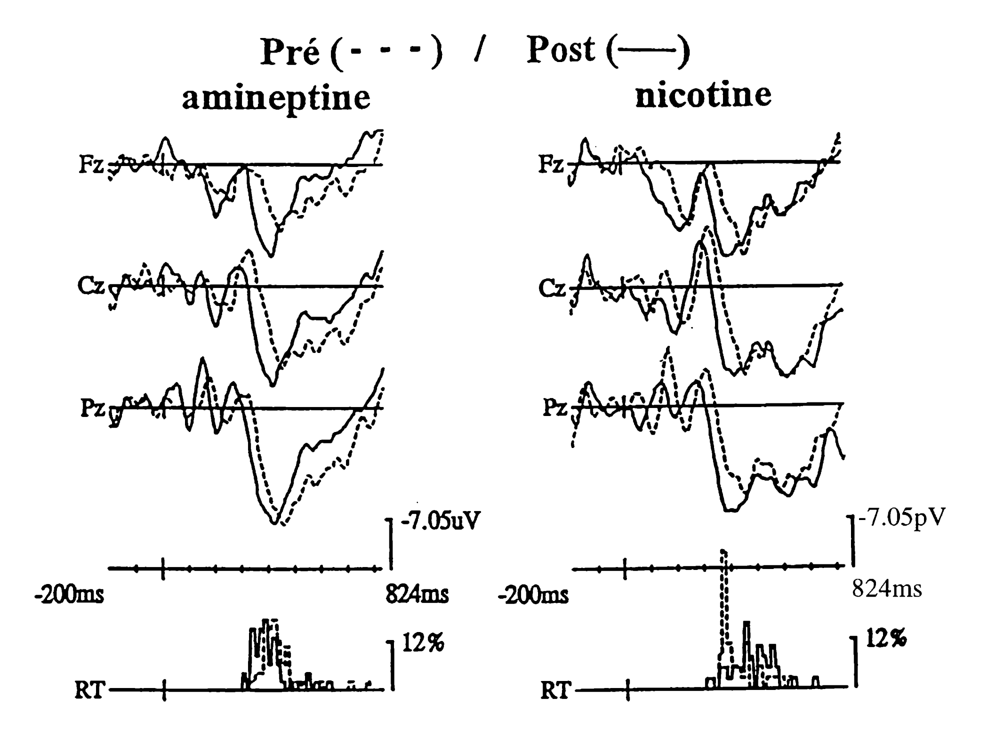


Fig. 1. *Tentative de mise en évidence d'un trait « réaction hédonique par activation corticale après stimulation des systèmes dopaminergiques par la nicotine ».*

Comparaison des effets pré-post (2 h d'intervalle) sur les potentiels évoqués cognitifs, de la consommation de tabac (3 petits cigares) et d'une prise unique d'amineptine (150 mg) chez un sujet. Les tracés étant très semblables, cela semble indiquer que la nicotine agit ici comme un agoniste dopaminergique (Jacques Le Houezec, communication personnelle).

EN GUISE DE CONCLUSION

L'heure n'est plus à un débat dualiste et l'environnement actuel des idées invite à une réflexion théorique, en vue d'une clinique compatible avec une véritable approche expérimentale, nous l'avons dit. Ce constat est aussi l'occasion de souligner combien il s'agit de mieux assumer la dualité d'une clinique dans laquelle la saisie sémiologique n'est pas la même lorsqu'elle obéit à des finalités théoriques et/ou de recherche ou bien au contraire à des fins thérapeutiques (Jouvent, 1990). C'est redire à nouveau qu'il n'y a pas des objets structuraux et des objets mouvants, il y a deux lectures, deux politiques d'appréhension du vivant : l'une se raccroche à des invariants, l'autre s'attache aux mouvements et aux facteurs de résistance aux changements; la personnalité, le tempérament, la nosographie, la physiologie, la psychopathologie sont autant de concepts et de disciplines qui se doivent d'assumer cette double lecture. La neurobiologie et les fortes découvertes qu'elle a pu réaliser récemment invitent à ne pas délaisser l'approche dynamique dans la pratique et la recherche sur les phénomènes mentaux.

Bibliographie

AKISKAL H. — An Integrative Perspective on Recurrent Mood Disorders : the Mediating Role of Personality. In : BECKER J. and KLEINMAN A. (Eds), *Psychosocial Aspects of Depression*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1990.

BARUCH P., FILTEAU M. J., BOUCHARD R., POURCIER E., VINCENT P., JOUVENT R. — The Dimensional Approach in Psychopharmacology : a Polysemous Concept. *J. Psychiatr. Neurosc.*, *1*, 43-56, 1992.

BUYDENS-BRANCHEY L., BRANCHEY M., FERGESON P., HUDSON J., MCKERNIN C. — Euphorogenic Properties of the Serotonergic Partial Agonist m-Chlorophénylpiperazine in Cocaine Addicts. *Arch. Gen. Psychiatry*, *50*, 1001, 1994.

ENGELBERG H. — Low Serum Cholesterol and Suicide. *Lancet*, *339*, 727-729, 1992.

EYSENCK H. J. — Drugs and Personality I. Theory and Methodology. *Journal of Mental Science*, *103*, 119-131, 1957.

HUTEAU M. — *Les conceptions cognitives de la personnalité*. PUF, Paris, 1985.

JOUVENT R. — Une clinique pour penser, une clinique pour agir. *Revue internationale de psychopathologie*, *2*, 291-295, 1990.

JOUVENT R., CARTON S. — L'émotion dérégulée. In WIDLÖCHER D., Traité de psychopathologie, PUF, Paris, 561-581, 1994.

JOUVENT R., PARTIOT A. — Les évolutions de la clinique. In : FEDIDA P., WIDLÖCHER D. *Les évolutions – Phylogénèse de l'individuation* eds. PUF, Paris, 1994.

KILZIEH N., CLONINGER R. — Psychophysiological Antecedents of Personality. *J. Personality Disorders*, spring sup. 100-117, 1993.

LE MOAL M., JOUVENT R. — Psychopathologie et modèles expérimentaux. *Psychologie française*, *4*, 210-216, 1991.

MENZA M. A., FORMAN N. E., SAGE J. J., GOLBE L. I. — Parkinson's Disease and Smoking : the Relationship to Personality. *Neuropsychiatry, Neuropsychology and Behaviral Neurology*, *6*, 4, 214-218, 1993.

PIAZZA P. V., DEMINIÈRE J. M., LE MOAL M., SIMON H. — Factors that Predict Individual Vulnerability to Amphetamines Self-administration. *Science*, *245*, 1511-1513, 1989.

PICHOT P., BOYER P., PULL C. B., REIN W., SIMON M., THIBAULT A. — Un questionnaire d'auto-évaluation de la symptomatologie dépressive, le questionnaire QD2. *Revue de psychologie appliquée*, *34*, 229-250, 1984.

PONS L., NURNBERGER J., MURPHY D., JOUVENT R., RIPS C., WIDLÖCHER D. — Mood-independent Aberrancies of Word Responses and Action of Lithium on their Repetition in Manic Depressed Illness. *Pharmacopsychiatry*, *20*, 227-229, 1987.

REUCHLIN M. — *Les différences individuelles dans le développement conatif de l'enfant*. PUF, Paris, 1990.

STRELAU J., ANGLEITNER A. (Eds) — *Explorations in Temperament : International Perspectives on Theory and Measurement*. Plenum Press, London, 1991.

THOMAS A., CHESS S. — *Temperament and Development*. Brunner-Mazel, New-York, 1977.

WIDLÖCHER D., JOUVENT R. — La personnalité et le tempérament : approches structurales ou dynamiques? *Revue internationale de psychopathologie*, *17*, 25-41, 1995.

8

INTELLIGENCE ARTIFICIELLE : UNE AIDE À LA RECHERCHE CLINIQUE EN PSYCHIATRIE?

par I. BICHINDARITZ

L'intelligence artificielle est une spécialité de l'informatique dont le but est d'étudier les moyens par lesquels les *ordinateurs* peuvent être rendus capables de *réaliser des tâches cognitives*, dans lesquelles *les êtres humains sont actuellement les meilleurs* (Shapiro, 1987). Par l'objet de son étude, l'intelligence artificielle fait partie intégrante des sciences cognitives et partage des intérêts communs avec les autres disciplines qui les composent. Cependant, la contrainte d'utiliser des machines lors de sa mise en œuvre la rattache prioritairement à l'informatique et aux sciences de l'information.

Le projet général de l'intelligence artificielle est aujourd'hui de concevoir et réaliser des systèmes computationnels capables d'aider et assister les humains dans la réalisation de tâches cognitives. Ces tâches peuvent être des tâches expertes, telles que des tâches de diagnostic, de conception ou de planification, ou des tâches de sens commun, telles que la compréhension du langage naturel, la reconnaissance d'une scène ou la détermination d'un chemin d'accès à un objet.

Dans la première partie de ce chapitre, les principes sous-jacents de l'intelligence artificielle sont explicités, en même temps que leur évolution. Dans la seconde partie, des systèmes intelligents sont présentés à titre d'exemple. Dans la troisième partie, des applications de l'intelligence artificielle en psychiatrie sont exposées. Dans la quatrième partie sont proposées des contributions potentielles de l'intelligence artificielle à la recherche en psychiatrie.

PRINCIPES D'INTELLIGENCE ARTIFICIELLE

Bien avant l'avènement du premier ordinateur, l'informatique a été rêvée comme un moyen de seconder l'être humain dans ses travaux intellectuels ou artistiques. Dès 1842, le mathématicien Babbage avait conçu, sans pouvoir la réaliser, la *machine analytique* comme un exécutant fidèle pour des calculs scientifiques complexes ou la composition de morceaux de musique (Hofstadter, 1985). En 1956, peu après l'invention des premiers ordinateurs (1930-1950), à la conférence de Dartmouth, un groupe de chercheurs passionnés d'informatique a défini le premier projet de l'intelligence artificielle, qui était de faire réaliser par une machine toutes les tâches cognitives de l'être humain. Les premiers systèmes d'intelligence artificielle ont été

d'abord appliqués aux domaines de prédilection des chercheurs, tels que les joueurs d'échec ou les démonstrateurs automatiques de théorèmes. Dans les années 60, des réseaux de neurones simples comme le perceptron, et les premiers systèmes experts ont été réalisés encore dans un but expérimental. Plus tard, dans les années 70, les premiers systèmes experts appliqués ont vu le jour, en particulier le célèbre Mycin dans un domaine médical (les infections bactériennes), et les premiers systèmes capables d'apprendre, tels qu'Arches et Eurisko. Plus récemment, l'essor des systèmes experts et des réseaux connexionnistes (successeurs des réseaux neuronaux) s'est intensifié, et de nombreuses applications de l'intelligence artificielle sont opérationnelles et commencent à répondre à son projet initial de réaliser les tâches cognitives de l'être humain (Winston, 1988). D'autres pôles de recherche rejoignent peu à peu l'intelligence artificielle, telles que les bases de données ou la réalité virtuelle. Un autre domaine important est celui de l'intelligence artificielle distribuée, chargée de gérer les ensembles d'ordinateurs travaillant en parallèle, éventuellement dans un but de coopération.

Au sein de l'informatique, l'intelligence artificielle a développé ses propres outils et méthodes, en particulier des langages de programmation tels que *Lisp*, *Prolog* ou le premier langage à objets, *Smalltalk* ainsi que des bases de données orientées objets.

Originellement, l'intelligence artificielle repose sur l'*hypothèse du calcul symbolique*. Celle-ci postule que les processus nécessaires à la production d'actions intelligentes peuvent être simulés par un ensemble de symboles physiques et un ensemble de mécanismes produisant, dans le temps, une série de structures construites à partir de ces symboles. Par exemple, les idées développées dans cet article sont exprimées par l'ensemble des symboles de l'écriture, et l'hypothèse du calcul symbolique y est ainsi réalisée. Des approches non symboliques sont utilisées en intelligence artificielle, tels que certains modèles connexionnistes, mais communiquent ou manipulent, à un moment ou à un autre, leurs résultats sous la forme de symboles.

Aujourd'hui, des tendances bien délimitées sont identifiables en intelligence artificielle, et la coopération est souvent tentée entre des techniques complémentaires. Les systèmes experts des années 60-70 ont pour formalisme de représentation des connaissances les règles de production, c'est-à-dire des unités de connaissance de petite taille du type *SI prémisses ALORS conclusions*. Ces règles, rassemblées dans une base de connaissances, sont couplées à un ensemble d'inférences représentées séparément dans un moteur d'inférences. La séparation entre connaissances et raisonnements est complète. D'autres systèmes, tels que par exemple les systèmes de raisonnement à partir de cas, ont exploré la voie de l'indissociabilité des connaissances et du raisonnement. Leurs bases de connaissances sont un ensemble de cas, pouvant être de grande taille. Enfin, les systèmes de raisonnement à partir de modèles expriment leurs connaissances sous la forme de modèles, ou ensembles interconnectés d'objets et de relations entre ces objets. Leur intérêt est dans la grande taille des unités de connaissance, donnant une vision globale d'un domaine. Les systèmes de raisonnement à partir de cas, tirant parti des connaissances expérimentales d'un domaine, sont avantageusement complémentés par les systèmes experts, et mieux encore par les systèmes de raisonnement à partir de modèles, du fait de la taille possiblement importante des unités de connaissance dans les deux formalisations.

Tous ces systèmes fonctionnent à un niveau symbolique, facilitant la communication au moyen de symboles, donc avec les êtres humains, par opposition aux modèles

dits connexionnistes où les unités de connaissance sont réduites à des neurones dits formels, éclatant davantage encore la connaissance que dans les systèmes experts. C'est alors l'extraction de cette connaissance qui devient difficile, car elle est située à un niveau non plus symbolique, mais sub-symbolique. Et surtout, l'abandon du niveau symbolique prive ces systèmes, véritables boîtes noires, de la capacité d'explication qui est la grande force des systèmes symboliques, tels que les systèmes experts, les systèmes de raisonnement à partir de cas, et les systèmes de raisonnement à partir de modèles. La complémentation entre ces approches est cependant une voie de recherche fructueuse.

Les systèmes à base d'algorithmes génétiques sont un autre type de tentative, celle de doter les systèmes artificiels de règles de génération de solutions potentielles et de sélection inspirées à l'origine par les règles issues de la génétique. Ils sont performants dans les assemblées de systèmes symboliques travaillant en parallèle.

La facilitation d'accès à des machines rapides et « puissantes » accélère la diffusion des méthodologies issues de l'intelligence artificielle. Cependant, dans cette discipline récente, l'innovation des chercheurs est encore, et sera peut-être toujours, le principal facteur de développement.

DES SYSTÈMES INTELLIGENTS

Dès le début de son développement, les chercheurs en intelligence artificielle se sont interrogés sur les limites de leur projet. En 1950, le mathématicien Alan Turing applique aux ordinateurs le théorème de Gödel, qui énonce l'impossibilité d'axiomatiser entièrement les mathématiques, et plus généralement les ressources de l'intelligence humaine. Néanmoins, il propose un test pour concrétiser le concept de machine capable de penser. Une machine (en fait un système tournant sur cette machine) peut penser si elle réussit à faire croire à un observateur humain qu'il communique avec une autre personne humaine, dans des conditions ne mettant pas en question son apparence « physique » : c'est le *test de Turing*. Ce test, utilisé en particulier pour valider un des premiers systèmes d'intelligence artificielle en psychiatrie, *Parry*, présente l'intérêt d'avoir entraîné un approfondissement de la notion d'intelligence chez une machine.

Une machine intelligente doit être capable de faire des découvertes. Le programme AM (Lenat, 1993) créé en 1978 propose une modélisation du processus de découverte par une machine dans le but de mieux appréhender, voire de démystifier, cette activité. Il est représentatif de la démarche de l'intelligence artificielle, qui procède en plusieurs étapes. Il faut d'abord choisir une activité cognitive, telle que la découverte. Puis développer des théories et des hypothèses sur cette activité, et l'implémenter sur un ordinateur. Ensuite, expérimenter le programme et chercher à trouver d'où vient l'intelligence restante dans l'activité modélisée. Les limites de l'intelligence des programmes ne sont souvent que le reflet des limites des connaissances de ses concepteurs. Pour AM, les tâches cognitives humaines peuvent être considérées comme des recherches dans un espace de solutions potentielles. Afin de guider les recherches et de les contraindre, un ensemble d'heuristiques adaptées à chaque situation est modélisé et formalisé, dans AM sous la forme de règles de production, associant des prémisses à des conclusions. Un exemple de règle heuristique simplifiée est que *SI des sous-ensembles extrêmes (tel que l'ensemble réduit à*

un élément) d'un problème existent, ALORS il peut être intéressant d'étudier comment ils ont été obtenus. AM a été appliqué avec succès aussi bien à des découvertes mathématiques (la découverte des nombres premiers, la multiplication) qu'à des découvertes de la vie quotidienne (le fil à couper le beurre). Programme pionnier dans son domaine, AM a permis de mettre en œuvre des concepts qui ont servi de base aux premiers systèmes experts.

En essayant de rendre les machines capables de tâches aussi complexes que la découverte scientifique, les chercheurs en intelligence artificielle ont compris le rôle primordial des connaissances dans l'intelligence, au point que certains assimilent l'intelligence à l'organisation judicieuse de vastes connaissances. L'importance des connaissances ne cesse d'augmenter au cours des progrès réalisés, à tel point qu'un des pôles prépondérants de recherche est constitué par l'acquisition des connaissances, considérée comme le goulot d'étranglement des systèmes artificiels. Plusieurs types de réponses lui sont apportés.

Dans le projet *Cyc* (Guha et Lenat, 1993), une tentative de rassembler des connaissances universelles est en cours de réalisation. Un enseignement de *Cyc* est que les connaissances les plus difficiles à formaliser et manipuler sont les connaissances de sens commun. En effet, celles-ci sont utilisées par les humains les plus simples, dans leurs pensées quotidiennes, mais sans être explicitées. Elles sont, de par leur généralité, largement applicables, et facilement transposables d'un domaine à un autre. Par contre, les connaissances expertes sont directement accessibles dans les encyclopédies, ou auprès des experts eux-mêmes, quel que soit leur domaine d'expertise. Néanmoins, elles sont dédiées à un domaine particulier, et difficilement transférables dans un autre domaine. De façon générale, les ordinateurs sont plus doués pour les disciplines expertes, tels que les mathématiques ou le jeu d'échec, que pour les disciplines de sens commun, tels que, pour un robot, s'orienter dans l'espace ou prendre un objet dans sa main. Ce système est construit sur deux niveaux, un niveau épistémologique permettant la communication, au moyen d'une sémantique simple et naturelle, d'un utilisateur quelconque avec les connaissances variées du système, et un niveau heuristique destiné à améliorer l'efficacité du raisonnement au moyen de procédures proches de la machine.

Une autre réponse est apportée par une nouvelle génération de systèmes capables d'apprendre de façon soit autonome, soit guidée par un expert : les systèmes d'apprentissage (Ganascia, 1991). L'apprentissage est un autre pôle de recherche particulièrement actif de l'intelligence artificielle. Il présente actuellement un ensemble très vaste de méthodes (Michalski, 1993), applicables à des domaines variés, pouvant aller de l'apprentissage sans connaissances préalables du domaine à l'apprentissage par amélioration de modèles du domaine (Clancey, 1992). Plusieurs stratégies sont envisagées, la synthèse ou l'analyse. Les inférences, ou raisonnements, utilisés peuvent être : la déduction, si l'inférence est une conséquence logique des prémisses; l'induction, si l'inférence est une généralisation des prémisses; ou l'analogie, si l'inférence repose sur un raisonnement par analogie.

Parmi les systèmes capables d'apprentissage, les systèmes de raisonnement à partir de cas utilisent principalement des inférences analogiques pour raisonner et pour apprendre (Kolodner, 1993). Dans ces systèmes, un cas est un ensemble particulier de données empiriques représentant une instance d'un type ou d'un concept. Un système de raisonnement à partir de cas est un système qui utilise, pour traiter un nouveau cas, un ou plusieurs cas précédemment rencontrés. Il tire avantage de l'idée qu'il est plus économique, chaque fois que cela est possible, de modifier un cas déjà

traité, que de réaliser un traitement totalement nouveau. Pour eux, chaque traitement réalisé doit être mémorisé s'il peut éviter, dans l'avenir, de recommencer le même traitement. Ces systèmes permettent de construire automatiquement une base de connaissances à partir d'un ensemble de cas jugé représentatif du domaine.

L'origine de ces systèmes remonte aux travaux (Schank, 1977) sur la compréhension du langage naturel, mais des applications dans des domaines médicaux commencent à être réalisées (Bichindaritz et Séroussi, 1992). En effet, cette méthodologie est bien adaptée à la médecine (Campbell, 1992), où les connaissances sont initialement théoriques, puis complémentées par des connaissances expérimentales sous la forme de cas de patients (Bichindaritz, 1994). La perspective de disposer de systèmes capables d'apprendre à chaque résolution d'un cas, et ainsi de devenir de plus en plus compétents, donc experts, est intéressante. Un autre intérêt de cette méthodologie est de proposer un modèle cognitif de l'intelligence prenant en compte l'étroitesse de ses liens avec la mémoire. D'autre part, elle permet aussi d'intégrer des connaissances théoriques du domaine.

Un exemple de système de raisonnement à partir de cas, appliqué au problème du diagnostic en audiologie clinique, est le système *Protos* (Porter *et al.*, 1993). Après avoir appris en interaction avec un expert à poser un diagnostic correct, sur 200 cas, représentant l'ensemble des cas rencontrés en une année par un clinicien débutant mais spécialisé dans le domaine, Protos est parvenu à un taux de réussite dans son diagnostic de 92 % à 100 %, soit mieux que le clinicien le plus expérimenté du domaine (l'expert).

Ainsi, présentant l'activité de recherche dans un espace plus ou moins vaste comme une activité sous-jacente aux tâches cognitives complexes, les systèmes intelligents utilisent leurs connaissances pour guider leur recherche et la faire converger vers une solution. Ils sont désormais capables d'acquérir ces connaissances de façon autonome et d'apprendre pour se perfectionner et s'adapter au changement. Un des principaux rôles de l'intelligence artificielle est donc de faire reculer la frontière entre le rationnel et l'irrationnel dans le raisonnement humain, donc de mieux le comprendre.

INTELLIGENCE ARTIFICIELLE EN PSYCHIATRIE

Les premiers systèmes d'intelligence artificielle réalisés en psychiatrie ont soulevé de nombreuses polémiques. Tantôt recherchés, en particulier par certains patients, tantôt décriés, pas seulement par les psychiatres, ils furent les jalons nécessaires sur le chemin de la maturité de l'intelligence artificielle. Des systèmes tels que *Doctor* et *Eliza* (Weizenbaum, 1976) des années 60, qui proposaient de simuler la conversation d'un psychiatre Rogérien avec un de ses patients, ou comme *Parry* (Colby, 1975) qui simulait la conversation d'un patient paranoïaque, étaient des programmes simplistes et ne méritaient pas l'intérêt trop grand qu'ils ont suscité.

Plus récemment, une modélisation connexionniste des déficits de saisie d'information dans la schizophrénie (Servan-Schreiber et Cohen, 1991) a tenté de modéliser les relations entre les déficits attentionnels et le traitement du contexte dans la schizophrénie. Reposant sur une analogie entre un réseau de neurones formels, désormais appelé modèle connexionniste, et une partie du cerveau humain, ici le cortex préfrontal, ce système stimule l'effet de la dopamine par une variation d'un

paramètre d'entrée des neurones formels. Sa validation a consisté à montrer la ressemblance entre l'altération des performances au test de Stroop d'un malade réel, et du système. L'intérêt de ces travaux est de proposer une modélisation possible de paramètres biologiques par un modèle artificiel, mais l'analogie entre les modèles connexionnistes et une partie du cerveau humain est elle aussi simpliste. Les systèmes de simulation de processus mentaux réalisés jusqu'à aujourd'hui reproduisent la modélisation de leur concepteur, et n'en sont que le reflet, certes sophistiqué.

Par contre, les systèmes experts en psychiatrie sont en plein essor, et tentent de suivre l'intérêt croissant des disciplines médicales pour ces systèmes. Le système *Adinfer* (Ohayon, 1993) propose une réalisation d'un système expert de diagnostic psychiatrique. Son auteur présente une revue des principaux systèmes experts en psychiatrie, et conclut à leur utilité à des fins pharmacologiques et épidémiologiques. Testé à une grande échelle, il pourrait servir d'« expert » commun pour les études multicentriques et augmenter la fiabilité entre les centres. Il apporte de même une réponse efficace au problème de la stabilité du diagnostic entre cliniciens (fidélité inter-juges) et pour un même clinicien dans le temps (fidélité test-retest). Ce système expert modélise le raisonnement de diagnostic psychiatrique, en envisageant plusieurs stratégies au cours du raisonnement (hypothético-déductif, par analogie...).

D'autres systèmes proches des systèmes experts sont les systèmes d'aide à la prise de décision diagnostique (Amaral *et al.*, 1992). L'accent est alors porté sur la flexibilité du raisonnement, obtenue par intégration de méthodes d'intelligence artificielle variées, parmi lesquelles les réseaux sémantiques, les règles de production, et la connexion avec des méthodes statistiques.

Les systèmes à venir, alliant les avancées de domaines de l'intelligence artificielle complémentaires, en particulier de l'apprentissage et de la communication homme-machine, proposeront des fonctionnalités encore plus adaptées à la psychiatrie.

RECHERCHE CLINIQUE EN PSYCHIATRIE

Parmi les applications potentielles de l'intelligence artificielle en psychiatrie, une voie privilégiée commence à s'ouvrir : l'aide à la recherche clinique. Il est plus facile à l'heure actuelle de parler des aides ponctuelles possibles à la recherche clinique en psychiatrie, mais il est déjà envisageable de concevoir un système intégré présentant une combinaison harmonieuse et cohérente des différentes propositions suivantes. D'autres domaines influençant la psychiatrie ont aussi des projets communs avec l'intelligence artificielle, tels que l'épidémiologie (Raschetti, 1992), et ces coopérations auront des conséquences, indirectes, sur la recherche clinique en psychiatrie.

Pour un spécialiste en intelligence artificielle, la psychiatrie est un domaine dit *faiblement théorisé*, par opposition aux domaines des sciences exactes, tels que les mathématiques ou la logique. C'est également un domaine où les informations sont de sources multiples et hétérogènes. Le grand nombre de données disponibles nécessite un choix pour ne garder que les variables les plus représentatives, en fonction de l'étude à réaliser.

La stratégie suivie par la recherche clinique en psychiatrie comporte plusieurs étapes (Guelfi, 1987). La première étape est de *définir l'hypothèse de départ*. Cette hypothèse doit être précise, concrètement vérifiable et assez simple. Une aide de l'intelligence au niveau du choix de l'hypothèse est de proposer des idées issues de ses résultats de recherche sur le raisonnement. Par exemple, une étude comparative entre des patients et des témoins des types de raisonnement (par analogie, induction ou déduction), de la représentation de leurs connaissances (classes, scripts, typologies de relations), de leurs heuristiques, du méta-raisonnement... En particulier, une application a été réalisée en psychiatrie pour étudier le raisonnement par analogie chez des schizophrènes, en essayant de déterminer quelle étape du raisonnement était plus spécifiquement perturbée.

Une autre possibilité pour cette première étape est de *constituer une base de connaissances* à partir des articles de la littérature concernant le domaine de recherche, et des domaines proches, pour pouvoir transposer, par raisonnement, par analogie, ou par apprentissage d'heuristiques, des hypothèses d'un domaine à un autre. Un système d'aide à la recherche en psychiatrie serait alors proche des systèmes d'aide à la découverte scientifique. Des travaux sur le traitement computationnel du langage médical ont déjà permis de réaliser automatiquement une base de connaissances extraite de la littérature d'un domaine médical (Sager *et al.*, 1987).

La deuxième étape de la recherche clinique en psychiatrie est la définition de la *nature des variables*. Les travaux sur le langage naturel permettent d'envisager d'inclure parmi les données de départ des informations textuelles issues de comptes-rendus médicaux, et ainsi d'élargir le choix possible des données étudiées. Dans cette étape, la principale difficulté est de neutraliser l'effet des variables dites incontrôlables. Les spécialistes en recherche clinique conseillent de choisir une population de référence aussi proche que possible de la population pathologique, et de considérer des groupes de grande taille. Un avantage de l'informatique est de pouvoir traiter des données à plus grande échelle que sans son concours, et l'intelligence artificielle accroît encore cette possibilité par rapport à l'informatique dite classique.

La troisième étape est le *recueil des données*. Sa réussite est conditionnée par la qualité des instruments d'évaluation (caractéristiques métrologiques optimales) et par des conditions de recueil optimales. Parmi les qualités métrologiques, les fidélités inter-juges, test-retest et inter-centres seraient augmentées par l'utilisation de systèmes experts, par exemple de diagnostic comme ceux présentés dans la section précédente. Les mêmes systèmes amélioreraient aussi les conditions de recueil par la neutralisation de l'interaction médecin-malade, sauf si celle-ci est l'objet de l'étude envisagée. D'autre part, les sous-domaines de l'intelligence artificielle tels que ceux spécialisés dans la communication homme/machine étudient comment faciliter cette communication et la rendre plus fiable et plus naturelle.

La quatrième étape est l'*analyse des résultats*. Traditionnellement en psychiatrie, elle est réalisée par des méthodes dites numériques, par référence aux méthodes de l'intelligence artificielle plutôt qualifiées de symboliques. Des projets communs sont en cours concernant les apports possibles de l'intelligence artificielle aux statistiques (Defays, 1989) et à l'analyse des données. L'intelligence artificielle peut leur apporter ses progrès dans le traitement de l'information. Ils permettront de rendre les logiciels de statistiques plus agréables à l'utilisateur, par la fourniture d'explications, de définitions et d'optimisations. Il s'agit d'incorporer dans les logiciels une partie des raisonnements des experts statisticiens. Les apports possibles touchent également

le recueil, comme indiqué précédemment, mais aussi la planification des expériences, le traitement des données et leur validation, l'identification des statistiques ou plus généralement des méthodes numériques adaptées. Ils traitent aussi de la possibilité d'élargir le choix des techniques statistiques, en étant plus proches de l'utilisateur et en le guidant pour éviter qu'il ne soit dépassé par la variété proposée. Kodratoff et Diday (1991) présentent des combinaisons potentielles entre analyses symbolique et numérique des données. En effet, l'analyse des données classiques étudie les données quantitatives ou qualitatives au moyen de méthodes numériques, d'outils mathématiques et statistiques. L'analyse symbolique des données classiques propose d'autres types de traitements, tels que construction de hiérarchies par généralisation (Bichindaritz, 1994) et héritage. Elle peut aider à l'interprétation des résultats d'analyses numériques. À l'autre extrémité, l'analyse numérique des données symboliques utilise des outils numériques classiques pour étudier les données symboliques. Elle est limitée par l'impossibilité de tirer avantage des connaissances du domaine. Enfin, l'analyse symbolique des données symboliques utilise les traitements symboliques cités précédemment.

Enfin, la cinquième étape est l'*interprétation des résultats*. Ici encore, l'aide d'un système intelligent serait de proposer des interprétations, par confrontation des connaissances du domaine avec les nouvelles connaissances, et par comparaison avec les connaissances de domaines différents. En particulier, le raisonnement à partir de cas et les méthodes d'apprentissage, à partir de modèles, sont spécialisés dans la production d'explications. Un avantage de l'intelligence artificielle par rapport à des méthodes numériques, et de manière évidente, une complémentarité, est d'expliquer les résultats obtenus (Bichindaritz, 1994).

Les atouts de l'intelligence artificielle sont de rendre possible un projet ayant une perspective d'intégration de ces propositions. Modélisant la recherche clinique en psychiatrie comme une tâche de découverte, le but d'un système à visée globale est de proposer un modèle cohérent, à partir d'un ensemble de cas de patients, multidimensionnel, et fidèle à la réalité clinique. Pour cela, un approfondissement des raisonnements des experts en recherche clinique, comme cela a été fait pour le raisonnement diagnostique, peut être tenté. Les résultats de projets comme *Cyc* apporteront des ensembles de connaissances et d'heuristiques générales, qu'il faudra compléter par des connaissances et des heuristiques propres du domaine.

CONCLUSION

Parmi les domaines d'application de l'intelligence artificielle, la psychiatrie a donné matière à peu de réalisations, en comparaison avec la médecine en général. D'ailleurs, ce n'est que depuis quelques années que les applications médicales de l'intelligence artificielle commencent à se développer à grande échelle, témoignant d'un retard par rapport à d'autres domaines, tels que la défense et l'industrie.

Si les systèmes experts et les systèmes de raisonnement à partir de cas vont rapidement devenir opérationnels pour l'aide au diagnostic en psychiatrie, les systèmes d'aide à la recherche clinique nécessitent un travail de modélisation plus complexe et à plus grande échelle, car touchant un nombre important de pôles de recherche tant en intelligence artificielle, qu'en statistiques, en analyse des données et dans d'autres disciplines encore.

En effet, le caractère hétérogène des données, particulièrement des données langagières ou à autres supports de communication, le degré faible de modélisation du domaine, et le caractère subjectif de l'activité clinique, rendent nécessaire la conception d'un système doué de connaissances importantes, tant sur le domaine, que de type heuristique, et de capacités d'apprentissage. Les travaux réalisés en intelligence artificielle préparent à des applications de cette envergure.

Cependant, la réussite d'une telle entreprise est conditionnée par l'intérêt des chercheurs cliniciens en psychiatrie qui participeront à ce projet. Ils en ressortiront sûrement plus confiants, s'ils ne l'étaient déjà, dans l'aide des machines pour décupler les facultés de raisonnement, de mémorisation et de communication humaines. D'autre part, le travail de formalisation et celui de modélisation nécessaires à la conception d'un système computationnel constituent de toute façon une avancée de la rationalisation du domaine d'application.

En tous les cas, les chercheurs en intelligence artificielle attendent avec impatience l'avènement des systèmes d'aide à la recherche en intelligence artificielle.

BIBLIOGRAPHIE

AMARAL M. B., SATOMURA Y., HONDA M., SATO T. — A Decision Support System using Natural Language Processing and the DSM-III-R. *MEDINFO, 92*, 526-530, 1992.

BICHINDARITZ I., SÉROUSSI B. — Contraindre l'analogie par la causalité. *Technique et science informatique, 11*, 4, 69-98, 1992.

BICHINDARITZ I. — A Case-Based Reasoning System Using a Control Case-Base. *Proceedings of the European Conference on Artificial Intelligence*, Amsterdam, 1994 (in press).

CAMPBELL J. A. — Computational Models of Case-Based Reasoning for Medicine. In : EVANS D. A., PATEL V. L., eds, *Advanced Models of Cognition for Medical Training and Practice*. Springer-Verlag, Berlin, 91-100, 1992.

CLANCEY W. J. — Model Construction Operators. *Artificial Intelligence, 53*, 1-115, 1992.

COLBY K. M. — *Artificial Paranoïa : Computer Simulation of Paranoid Processes*. Elmsford, New York, Pergamon Press, 1975.

DEFAYS D. — Statistics and Artificial Intelligence. In : DIDAY E., eds, *Data Analysis, Learning Symbolic and Numeric Knowledge*. Nova Science Publishers Inc., New York, 381-385, 1989.

GANASCIA J. G. — *L'âme-machine : les enjeux de l'intelligence artificielle*. Seuil, Paris, 1990.

GUELFI J. D., BOYER P., CONSOLI S., OLIVIER-MARTIN R. — *Psychiatrie*. Presses Universitaires de France, Paris, 1987.

GUHA R. V., LENAT D. B. — Cyc : a Midterm Report. In : BUCHANAN B. G., WILKINS D. C., eds, *Readings in Knowledge Acquisition and Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 839-866, 1993.

HOFSTADTER D. — *Gödel, Escher, Bach : les brins d'une guirlande éternelle*. InterEditions, Paris, 1985.

JOHNSON-LAIRD P. N. — *Human and Machine Thinking*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1993.

KODRATOFF Y., DIDAY E. — *Induction symbolique et numérique à partir des données*. Vol. 1, CEPADUES-Éditions, Toulouse, 1991.

KOLODNER J. L. — *Case-Based Reasoning*. Morgan Kaufmann Publishers, San Mateo, CA, 1993.

LENAT D. B. — The Ubiquity of Discovery. In : BUCHANAN B. G., WILKINS D. C., eds, *Readings in Knowledge Acquisition and Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 797-811, 1993.

MICHALSKI R. S. — Toward a Unified Theory of Learning : Multistrategy Task-Adaptive Learning. In : BUCHANAN B. G., WILKINS D. C., eds, *Readings in Knowledge Acquisition and Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 7-38, 1993.

OHAYON M. — Utilisation des systèmes experts en psychiatrie. *Revue can. de psychiatrie*, *38*, 203-211, 1993.

PORTER B. W., BAREISS R., HOLTE R. C. — Concept Learning and Heuristic Classification in Weak-Theory Domains. In : BUCHANAN B. G., WILKINS D. C., eds, *Readings in Knowledge Acquisition and Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 741-758, 1993.

RASCHETTI R. — EPIAIM : Artificial Intelligence and Epidemiology. In : NOOTHOVEN VAN GOOR J., CHRISTENSEN J. P., eds, *Advances in Medical Informatics*. IOS Press, Amsterdam, 102-107, 1992.

SAGER N., FRIEDMAN C., LYMAN M. S. — *Medical Language Processing*. Addison-Wesley, Reading, MA, 1987.

SCHANK R. C. — *Scripts, Plans, Goals and Understanding*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1977.

SHAPIRO S. C. — *Encyclopedia of Artificial Intelligence*. Wiley Interscience, 1987.

SERVAN-SCHREIBER D., COHEN J. D. — Modèles de neuromodulation et déficits de saisie d'information dans la schizophrénie. *Revue Internationale de Psychopathologie*, *3*, 113-134, 1991.

SHORTLIFFE E. H. — The Adolescence of AI in Medicine : Will the Field Come of Age in the '90s? *Artificial Intelligence in Medicine*, *5*, 93-106, 1993.

WEIZENBAUM J. — *Computer Power and Human Reason*. Academic Press, San Francisco, CA, 1976.

WINSTON P. H. — *Intelligence artificielle*. InterEditions, Paris, 1988.

9

MMPI et MMPI-2

par J. D. Guelfi, M. Simon et I. Gillet[1]

HISTORIQUE

L'inventaire multiphasique de personnalité du Minnesota (MMPI) a été créé en 1943 par un psychologue; S. R. Hathaway, et un psychiatre, J. C. McKinley. Il explore les aspects caractériels et affectifs de la personnalité. Les échelles constituant ce questionnaire ont été établies par groupes contrastes entre, d'une part, une population générale et, d'autre part, des groupes pathologiques provenant, pour la plupart, du service neuropsychiatrique des Hôpitaux de l'université de Minnesota. Le MMPI, dans sa forme initiale, était constitué de 550 items, recouvrant un champ assez vaste, depuis la condition physique jusqu'aux attitudes morales ou les conduites sociales d'un individu.

Pour chaque item, le sujet doit répondre «vrai», «faux» ou «je ne sais pas». Les items sont répartis en 4 échelles de validité (ou attitudes de réponses) et 10 échelles cliniques. De plus, 10 sous-échelles regroupées en O (*obvious*, «signification évidente»), et S (*subtle*, «signification cachée») ont été développées par Wiener à partir des items de 5 des échelles cliniques. Par la suite, Harris et Lingoes ont construit 28 sous-échelles complémentaires, illustrant les différentes facettes de 6 échelles cliniques. De nombreuses échelles additionnelles et divers indices ont été élaborés à partir des 550 items de base comme l'Échelle d'anxiété manifeste de Taylor, l'Échelle de force du Moi de Barron, l'Indice de sincérité de Gough ou l'Indice d'anxiété deWelsh.

DESCRIPTION DES ÉCHELLES

Les quatre échelles d'attitudes de réponses permettent de pondérer, de corriger, voire d'invalider les résultats obtenus sur les échelles cliniques; elles peuvent mettre en évidence les falsifications conscientes ou inconscientes des protocoles.

1. Avec la collaboration, pour la première étude française confrontant les données du compte-rendu informatisé du MMPI-2 avec les jugements cliniques des psychiatres traitants, les Drs : Aubin, Boyer, Corcos, Dardennes, Dhote, Doublet, Duneton, Foulon, Guelfi, Guillibert, Hakim-Kreis, Hassan, Heim, Kalck, Léonard, Lévy-Garboua, Matheron, Olivier-Martin, Seror, Thibault, Vautherin et Waintraub, et avec la participation des psychologues qui ont assuré la passation des MMPI et contribué au dépouillement des avis des cliniciens sur les comptes-rendus informatisés, Mmes A. Brun, C. Monier, F. Seunevel et L. Svarna.

– **Échelle?** (nombre de réponses «*je ne sais pas*») : elle indique l'indécision, le doute, la réticence, voire la falsification.

– **Échelle L** (*lie*, mensonge) : elle explore la tendance, consciente ou non, à se présenter sous un jour systématiquement favorable; elle permet éventuellement de réajuster les notes des échelles cliniques.

– **Échelle F** (échelle dite «d'anomalie psychique») : elle permet de contrôler la validité de l'ensemble du test; elle peut refléter soit l'incompréhension des questions, soit le fait de répondre au hasard; mais, le plus souvent, elle indique des réponses rares, déviantes.

– **Échelle K** : elle évalue l'attitude du sujet vis-à-vis du test ou de ses symptômes. Les notes extrêmes indiquent soit la surestimation soit la sous-estimation systématique des symptômes. L'échelle K sert à corriger si besoin la note des échelles cliniques **Hs, Pd, Pt, Sc et Pa**.

– **Les échelles cliniques** permettent de mesurer 10 dimensions empruntées à la nosologie psychiatrique. La liste indique les significations cliniques les plus fréquentes des notes élevées sauf pour l'échelle Mf, pour laquelle une note basse peut aussi être interprétée.

– **Hs** (hypocondrie) : inquiétude exagérée concernant son état de santé, tendance à se plaindre perpétuellement de maux et de troubles.

– **D** (dépression) : sentiment d'inutilité et d'incapacité à envisager l'avenir, tendance à l'inquiétude.

– **Hy** (hystérie) : immaturité, revendication affective, manipulation d'autrui, ou symptômes de type «conversion».

– **Pd** (déviation psychopathique) : égocentrisme, mépris des normes sociales, absence de réactions émotionnelles profondes et incapacité de tirer bénéfice de l'expérience acquise.

– **Mf** (masculinité-féminité) : intérêts et attitudes rapprochant le sujet du sexe opposé, surtout chez les hommes (note basse : conformité exagérée aux stéréotypes de son sexe).

– **Pa** (paranoïa) : méfiance, hypersensibilité, idées de persécution avec ou sans idées de grandeur. Une note anormalement basse caractérise parfois un sujet paranoïaque hypercontrôlé qui parvient à ne pas se trahir directement.

– **Pt** (psychasthénie) : doute, indécision, fatigabilité.

– **Sc** (schizophrénie) : groupement hétérogène portant sur des pensées ou des comportements bizarres et inhabituels comme le repli sur soi ou une vie subjective séparée de la réalité.

– **Ma** (hypomanie) : hyperactivité de pensée et d'action.

– **Si** (introversion sociale) : évitement des contacts sociaux.

APPLICATIONS ET PREMIÈRE ADAPTATION FRANÇAISE

Le MMPI est à l'heure actuelle le questionnaire de personnalité le plus utilisé au monde, tant dans la pratique clinique, comme aide au diagnostic, que dans les travaux scientifiques qu'il suscite. De nombreuses adaptations dans des cultures et des langues différentes ont été réalisées, qui ont permis de confirmer la validité de cet

instrument. La première adaptation française du MMPI a été publiée en 1959 par P. Pichot et J. Perse (voir S. R. Hathaway et J. C. McKinley, version française, Éditions du Centre de psychologie appliquée). Les normes françaises ont été construites sur un échantillon représentatif de 321 sujets (146 hommes et 175 femmes). La comparaison du profil moyen de l'échantillon français avec l'échantillon américain a mis en évidence quelques différences significatives entre les deux populations (les Français obtenant, en moyenne, des notes plus élevées sur les échelles L, F, D, Sc, Si pour les deux sexes, Mf pour les hommes et Hy pour les femmes).

LE MMPI-2

La révision du MMPI a été entreprise à partir de 1980 par James N. Butcher et a été publiée en 1989 avec un échantillon de 2600 personnes (1462 femmes et 1138 hommes). Cette nouvelle version, fondée sur les mêmes principes de base, a eu pour objectif de retravailler la formulation de certains items (ambigus, désuets ou sexistes), de créer des items couvrant de nouveaux champs et enfin d'établir de nouvelles normes nationales. Cette forme contient 567 items. Une recherche internationale (24 pays) a eu pour but de garantir la validité transculturelle de l'inventaire tout en respectant les différences nationales. Ceci n'avait pas été réalisé pour le MMPI original (c'est-à-dire, aux États-Unis, au sein des diverses « minorités »). Enfin, plusieurs progrès ont aussi paru nécessaires sur le plan purement psychométrique, avec de nouvelles procédures d'établissement des notes T uniformes (m = 50; ET = 10). Ces notes ont une signification clinique dès 65 (et non plus 70 comme dans le MMPI). Des indicateurs additionnels de « validité », ou plutôt, d'attitudes vis-à-vis de l'épreuve ont été créés ainsi que diverses échelles (appelées échelles de contenu) dont l'unidimensionnalité, ou du moins l'homogénéité, a été vérifiée par des coefficients alpha de Cronbach.

Les échelles d'attitudes

Outre les échelles classiques : L (mensonge) de 15 items, F de 60 items, K de 30 items et « ? » (rendant le profil invalide à partir de la note de 30) et la validité « incertaine » de 11 à 29 selon J. Butcher, les échelles additionnelles d'attitudes s'intitulent « Back F Scale ou Fb », « TRIN » et « VRIN ». L'échelle « Back F » est constituée de 40 items situés dans la deuxième moitié du test. Elle permet de voir si l'attitude du sujet a changé au cours même de l'examen.

L'échelle TRIN ou *True Response Inconsistency Scale* est constituée de paires d'items à contenu opposé et indique la tendance à donner un type de réponses systématique quel que soit le contenu des items. Ainsi, certains sujets donnent-ils l'impression d'être toujours « d'accord » avec les propositions (notes élevées) ou, à l'inverse, d'être toujours « contre » (notes basses). L'échelle VRIN ou *Variable Response Inconsistency Scale* (Tellengen, 1988) est constituée de paires d'items à contenu similaire ou opposé. Une note élevée indique une tendance générale à répondre de façon inconsistante ou contradictoire aux items.

Les principales échelles cliniques

Certaines des dix échelles peuvent être subdivisées en sous-échelles qui explorent divers constituants d'une dimension principale.

Tableau 1. MMPI-2

Échelles principales			*Nombre d'items*	*Sous-échelles**
1	HS	Hypocondrie	32	H-L ou O-S
2	D	Dépression	57	5 H-L**
3	Hy	Hystérie	60	5 H-L***
4	Pd	Psychopathie	50	5 H-L + O-S
5	Mf	Masculinité/ Féminité	56	
6	Pa	Paranoïa	40	3 H-L + 2 O-S
7	Pt	Psychasthénie	48	
8	Sc	Schizophrénie	78	6 H-L
9	Ma	Hypomanie	46	4 H-L + O-S
10	Si	Introversion sociale	69	3 H-L + 3****

* O-S = *Obvious* et *Subtle scales* de Wiener et Harmon dont la validité reste discutée (J. Butcher).
H-L = sous-échelles de Harris et Lingoes.

** *Subjective depression, psychomotor retardation, physical malfunctioning, mental dullness, brooding.*

*** *Denial of social anxiety, need for affection,* lassitude-malaise, *somatic complaints, inhibition of aggression.*

**** 3 sous-échelles d'introversion sociale de Ben-Porath, Hosketler, Butcher and Graham (*shyness and self-consciousness, social avoidance and alienation-self and others*).

Les échelles additionnelles

Nombre de ces échelles existaient déjà dans la version initiale du MMPI. La plupart n'avaient pas fait l'objet d'études de validation en France. Elles figurent ci-dessous avec l'année de leur premier développement et le nombre d'items qui les constituent dans le MMPI-2. Figurent en outre de nouvelles échelles additionnelles dont la validation aux USA est récente.

Tableau 2. *Origines des échelles additionnelles*

Intitulé	Année	Nombre d'items
Échelles additionnelles		
Anxiété de Welsh (A)	1956	39
Refoulement de Welsh (R : Repression)	1956	37
Force du Moi (ES : Ego Strength) (Équilibre, adaptation au stress et sentiment d'adéquation, Barron)	1953	52
McAndrew Révisée (MAC-R) *Mac Andrew Alcoholism Scale Revised* (tendance à developper une addiction)	1965	49
(*Overcontrolled Hostility*, OH) Agressivité hypercontrôlée Megargee *et al.*	1967	28
Dominance de Gough *et al.* (Do)	1951	25
Responsabilité sociale de Gough (RE)	1952	30
Inadaptation scolaire (Kleinuntz)	1961	41
Nouvelles échelles additionnelles		
Post Traumatic Stress Disorders Scales	1984	46
Keane *et al.* (PK) et PS de Schlenger et Kulka	1987	60
Marital Distress Scale (MDS) de Huemboe *et al.*	in press	14
Addiction Potential Scale (APS) de J. Butcher	in press	39
Addiction Admission Scale (AAS) de Greene, Deweed, Butcher et Ben-Porath	1991	13

Les échelles de contenu (Content scales)

Quinze échelles originales se sont avérées homogènes. Certaines d'entre elles ont néanmoins été subdivisées en sous-échelles expérimentales faisant l'objet de développements actuels par Ben-Porath et Sherwood. Elles sont indiquées ci-dessous entre parenthèses.

ANX *Anxiety*, anxiété, 23 items.

FRS *Fears*, peurs, 23 items.

OBS *Obsessiveness*, obsessionnalité, 16 items.

DEP *Depression*, dépression (manque de motivation, dysphorie, autodépréciation et idéation suicidaire), 33 items.

HEA *Health Concerns*, préoccupations concernant la santé (symptômes gastro-intestinaux, symptômes neurologiques et préoccupations concernant la santé en général), 36 items.

BIZ *Bizarre Mentation*, idéation bizarre (symptomatologie psychotique et caractéristiques schizotypiques), 24 items.

ANG *Anger*, colère (comportement explosif et irritabilité), 16 items.

CYN *Cynism*, cynisme (misanthropie et relations suspicieuses à autrui), 23 items.

ASP *Antisocial Practices Attitudes* (attitudes antisociales et comportement antisocial), 22 items.

TPA Type A (impatience et compétitivité), 19 items.

LSE *Low Self-Esteem*, faible estime de soi, manque de confiance en soi, sentiment d'inutilité), 24 items.

SOD *Social Discomfort*, malaise social (introversion, timidité), 24 items.

FAM *Family Problems*, problèmes familiaux (conflits familiaux et désaffection familiale), 25 items.

WRK *Work Interference*, attitudes susceptibles de perturber le travail, 33 items.

TRT *Negative Treatment Indicators*, indicateurs d'un mauvais résultat thérapeutique (motivation faible et incapacité à se dévoiler ou croyance que personne ne peut aider ou comprendre), 26 items.

MMPI-2 ET ÉTUDES MÉTROLOGIQUES

Les études américaines

La consistance interne

Les coefficients de Cronbach varient pour les échelles cliniques de .34 à .85 chez les hommes ; ils ne sont inférieurs à .50 que pour l'échelle Pa (.34) ; ils varient de .37 à .87 chez les femmes ; ils sont inférieurs à .50 pour deux échelles : Mf (.37) et Pa (.39). Pour les échelles additionnelles, les coefficients alpha publiés par J. Butcher *et al.*, en 1989, varient de .89 (anxiété) à .34 chez les hommes, une seule échelle se caractérisant par un coefficient inférieur à .50 : l'échelle O-H (*Overcontrolled Hostility*). Ils varient de .91 (PS) à .24 chez les femmes, deux échelles se caractérisant par des coefficients inférieurs à .50 : l'échelle Mac Andrew révisée (.45) et l'échelle O-H (.24). Pour les échelles de contenu, les coefficients varient de .68 (type A chez les femmes) à .86 (cynisme chez les hommes).

L'analyse factorielle

L'analyse factorielle des réponses des 1 138 sujets masculins réalisée sur les 10 échelles cliniques et 3 échelles d'attitudes (L, F et K), avec rotation varimax, montre 4 facteurs. Un premier facteur de trouble général de l'adaptation (*General maladjustment*) extrait 35,8 % de la variance. Les saturations les plus élevées dans ce facteur concernent les échelles Sc, F, Pt, Hs et Pd. Le facteur II (16 % de la variance) est essentiellement constitué de deux échelles : Hy et K. L'auteur du manuel, J. Butcher, propose de l'intituler hypercontrôle (*overcontrol*). Le facteur III, d'introversion sociale (12,6 % de la variance), regroupe Si, D (saturations positives) et Ma (saturation négative). Enfin, le facteur IV, de masculinité-féminité (8,5 % de la variance) est représenté par une échelle unique : Mf. Ces mêmes quatre facteurs sont retrouvés dans l'échantillon féminin.

Études françaises

Formulation française des items du MMPI-2

La mise au point de la version française des items du MMPI-2 a été réalisée par un comité d'experts (P. Pichot, J. Perse et deux psychologues des Éditions du Centre de psychologie appliquée (ECPA)). Une première traduction des items nouveaux ou modifiés (N = 107) a été effectuée par deux psychologues traductrices bilingues. La confrontation des versions a abouti à la mise au point d'une formulation consensuelle pour chacun des items. Ce premier pool d'items a ensuite été soumis à une rétro-traduction (*back-translation*) effectuée par deux autres psychologues bilingues. La comparaison des rétro-traductions avec la version américaine d'origine a permis, après analyse, de stabiliser le contenu de ces items. De plus, les items français issus de la première version ont tous été étudiés de nouveau dans le but d'actualiser ou de simplifier, si nécessaire, leur formulation. 180 items ont ainsi été réaménagés en fonction de l'évolution du langage parlé de la population française.

Expérimentation

La première forme expérimentale de la version française du MMPI-2 a été administrée à 259 sujets « tout venant » (92 hommes et 167 femmes) de janvier à octobre 1993.

Résultats

Analyse des pourcentages de réponses : la comparaison des pourcentages de réponses semblables « *Vrai* » (IEF ou *Item Endorsement Frequency*) entre les échantillons français et américains a permis de classer les items selon leur taux de concordance (tableau 3).

45 items (7,9 %) pour les hommes et 48 items (8,5 %) pour les femmes présentent une différence de plus de 25 %. Cette différence est considérée comme significative par le département du Minnesota de J. Butcher. Elle a conduit à examiner de nouveau les items en question, tant du point de vue de la précision et de la qualité de la traduction, que du point de vue de la pertinence (validité de contenu) des items pour la population française. Dans la double perspective : de conserver l'équivalence interculturelle et de prendre en compte les différences culturelles et linguistiques, certains items ont été modifiés. Pour cela, il a parfois été nécessaire de s'éloigner de la traduction littérale du texte américain pour « tirer le sens » de l'item français vers le contenu de son échelle d'appartenance.

Tableau 3. *Les différences de pourcentages (en valeur absolue) observées sur les 567 items expérimentaux*

% de différence	*Hommes*	*Femmes*
26 et +	45 items	48 items
22-25	21 items	37 items
16-21	72 items	64 items
10-15	130 items	105 items
0-9	299 items	313 items

Analyse factorielle

Une analyse factorielle en composantes principales avec rotation Varimax a été effectuée séparément sur les données des hommes et des femmes. La structure mise en évidence sur les échantillons français est très comparable à celle obtenue aux États-Unis ainsi que dans d'autres pays.

Quatre facteurs isolés (78 % de la variance totale) peuvent être interprétés de la façon suivante : trouble général de l'adaptation; introversion sociale; hypercontrôle; masculinité-féminité. Ces résultats devront évidemment être confirmés sur des échantillons plus larges.

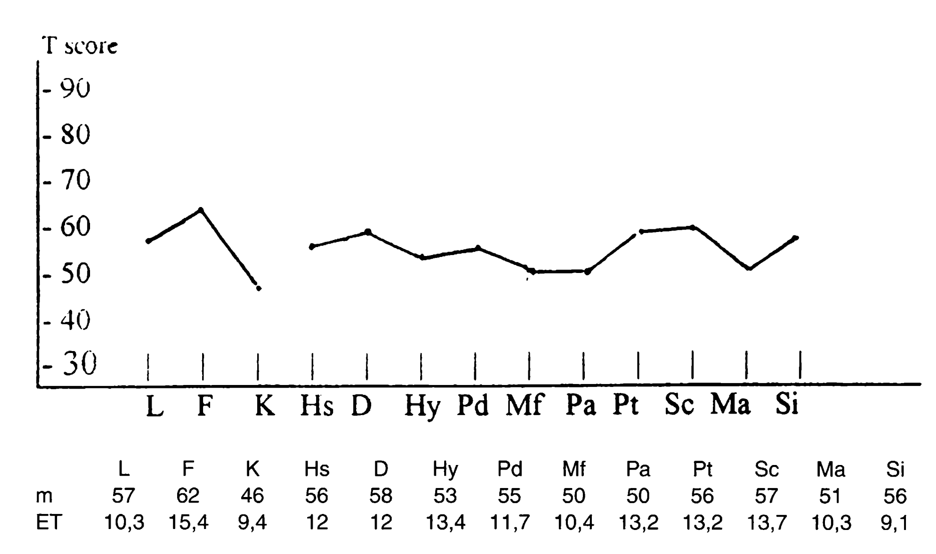
Une analyse en composantes avec rotation Varimax a été réalisée sur les échelles de contenu. La structure obtenue dans l'échantillon féminin est constituée de deux facteurs. Le premier est constitué des échelles WRK, LSE, DEP, TRT, ANX, SOD, OBS, FRS et HEALTH, échelles de symptômes d'allure névrotique et de troubles de l'adaptation émotionnelle ou sociale. Le facteur 2 (TPA, CYN, ASP, ANG, FAM, BIZ) semble plus constitué de variables comportementales ou de conflits dans les relations interpersonnelles.

Analyse du profil moyen des MMPI-2

Les profils moyens des échantillons français ont été construits d'après les normes américaines (note T : moyenne = 50; écart-type = 10). Les graphiques 1 et 2 illustrent les résultats obtenus. L'échelle F est significativement plus élevée en France (1 écart-type). D'autres échelles sont légèrement plus élevées en France (entre 1/2 et 1 écart-type) comme L, Hs, D, Pt, Sc, Si pour les deux sexes, ainsi que Mf chez les femmes.

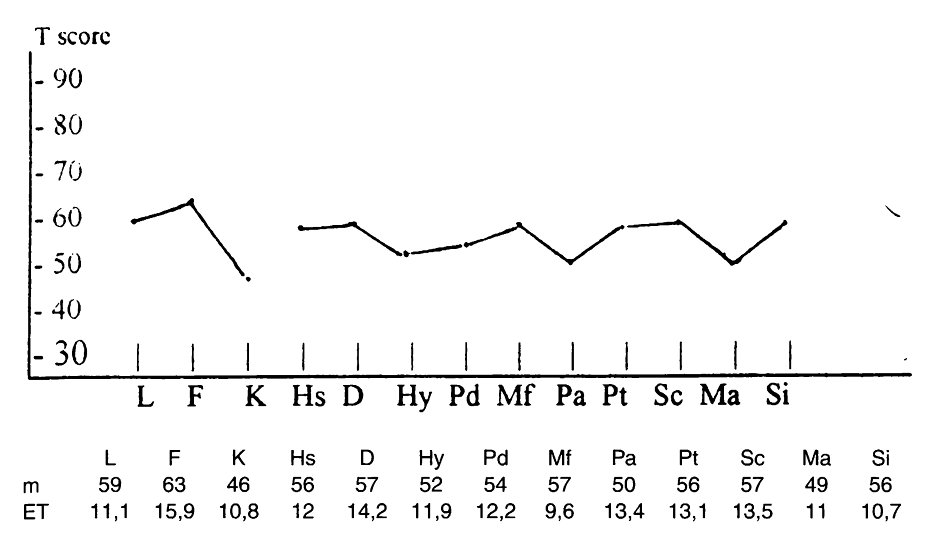
Étude Test-Retest

Une étude test-retest avec 12 sujets bilingues n'a pas montré de différences significatives des réponses entre les deux groupes.



	L	F	K	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	Si
m	57	62	46	56	58	53	55	50	50	56	57	51	56
ET	10,3	15,4	9,4	12	12	13,4	11,7	10,4	13,2	13,2	13,7	10,3	9,1

Graphique 1. *Profil moyen de l'échantillon français masculin N = 92*



	L	F	K	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	Si
m	59	63	46	56	57	52	54	57	50	56	57	49	56
ET	11,1	15,9	10,8	12	14,2	11,9	12,2	9,6	13,4	13,1	13,5	11	10,7

Graphique 2. *Profil moyen de l'échantillon français féminin N = 167*

DISCUSSION

Les items présentant plus de 25 % de différences de réponses entre échantillons français et américain ont été systématiquement revus. Le plus souvent, de simples modifications comme le renversement de la phrase ont suffi. Néanmoins, des difficultés relativement systématiques sont apparues, par exemple avec l'emploi du conditionnel ou des verbes de fréquence ou d'intensité : le conditionnel paraissant plus éloigné de la réalité actuelle pour les Français et plus proche d'une éventualité probable pour les Américains. Par ailleurs, l'aspect quantitatif de l'adverbe a semblé présenter un poids différent dans les deux langues et il a parfois été nécessaire de renforcer le sens de certains items en proposant des verbes moins hypothétiques et plus catégoriques.

Pour certains items, il a fallu restituer le contexte de l'item et rechercher des équivalences culturelles. Par exemple, l'item 384 «*I liked playing 'house' when I was a child*» aurait dû être traduit par «*jouer au papa et à la maman*». Cette expression a été jugée comme ayant une trop forte connotation sexuelle (dépassant la simple identification au rôle). La traduction choisie a été «*jouer à la dînette*».

Les principaux problèmes de traduction ayant été réglés, des différences entre les profils moyens semblent refléter d'authentiques différences culturelles. Les Français reconnaissent plus facilement chez eux des éléments de vulnérabilité, de faiblesse, d'impulsivité, de crainte des maladies, etc. Ils tendent à privilégier l'individu par rapport au groupe. Les Américains sont plus ouvertement extravertis, valorisant les règles sociales et privilégiant les activités collectives.

MMPI-2 ET CONFRONTATION DES COMPTES-RENDUS INFORMATISÉS AVEC LES JUGEMENTS CLINIQUES DE PSYCHIATRES TRAITANTS

Cette étude préliminaire a consisté à analyser les jugements des psychiatres traitants sur les rapports informatisés des données en provenance de 100 MMPI-2 auxquels des patients suivis en psychiatrie avaient répondu entre janvier et juin 1993. L'enquête a eu lieu dans deux départements spécialisés de psychiatrie à Paris : hôpital Sainte-Anne, Clinique des maladies mentales et de l'encéphale : CMME (Pr Samuel-Lajeunesse) et hôpital international de l'université de Paris (Pr Ph. Jeammet). Chaque psychiatre traitant devait, après lecture du rapport informatisé rédigé dans le département de J. Butcher, donner son opinion sur la qualité du texte proposé (en 5 degrés) en distinguant 5 sections différentes du rapport. Ainsi était-il demandé si l'information fournie dans chaque section du rapport était insuffisante (A), ou se résumait à « quelques informations » (B) ou était « suffisante » (C), « plus que suffisante » (D) ou « approfondie » (E). Les différentes sections considérées étaient : la validité du rapport, le tableau symptomatique, la description des relations interpersonnelles, les considérations diagnostiques et les considérations thérapeutiques du rapport.

Ensuite, une appréciation globale était demandée : « quel pourcentage des affirmations contenues dans le rapport apparaît comme pertinent pour décrire le patient en question : moins de 20 % (A), de 20 à 39 % (B), de 40 à 59 % (C), de 60 à 79 % (D) et de 80 à 100 % (E). Enfin, le diagnostic clinique le plus vraisemblable était demandé, établi sur les bases de la classification américaine du DSM-III-R ou de la classification internationale de l'OMS (CIM-9 ou CIM-10).

L'enquête a porté sur 100 patients suivis, hospitalisés (N = 54) ou ambulatoires (N = 46).

Les psychiatres ayant évalué les rapports informatisés des 100 MMPI-2 ont été au nombre de 22. Leur nombre d'années d'expérience clinique varie : de 4 à 6 ans (N = 5 médecins pour 33 MMPI-2 analysés), de 6 à 10 ans (N = 7 médecins pour 33 MMPI-2) et de plus de 10 années d'exercice (N = 10 médecins pour 34 MMPI-2).

Le nombre d'heures passées avec chacun des patients dans l'enquête a varié de 2 heures à plus de 6 heures pour la grande majorité d'entre eux.

Population étudiée

La population de l'étude a comporté 46 hommes et 54 femmes d'âge moyen 34 ans (écart-type 16). Les trois-quarts des patients avaient entre 20 et 40 ans.

Leurs niveaux socio-culturels appréciés selon le nombre d'années d'études étaient les suivants :

moins de dix années	N = 21
de 10 à 12 années	N = 26
de 13 à 14 années	N = 23
15 années et plus	N = 30

Les diagnostics cliniques ont été regroupés selon les axes et les principales classes du DSM-III-R.

Trouble isolé de l'axe I, psychotique	N = 30
Trouble isolé de l'axe I, dépressif ou anxieux	N = 19
Trouble isolé de l'axe I, des conduites alimentaires	N = 01
Autres troubles de l'axe I, somatoforme ou obsessionnel	N = 02
Troubles isolés de l'axe II, trouble de la personnalité	N = 25
Troubles mixtes : trouble de la personnalité + 1 trouble de l'axe I	N = 22
Trouble mixte, somatoforme + trouble des conduites alimentaires	N = 01

Au total : 47 patients sur 100 étaient considérés comme atteints d'un trouble de la personnalité. Chez 25 d'entre eux, le trouble paraissait isolé; chez les 22 autres, un trouble de l'axe I était associé au trouble de la personnalité : trouble affectif (N = 11) et trouble des conduites alimentaires (N = 11).

Les attitudes de réponse des 100 patients ont été les suivantes :

Coopérants (profils valides)	N = 20
« Vertueux » (*Virtuous*) se présentant sous un jour extrêmement favorable (profils valides)	N = 25
Exagération-dramatisation (profils valides)	N = 37
Inconsistance des réponses (conduisant dans la quasi-totalité des cas à un profil invalide selon les résultats aux échelles TRIN, VRIN ou Fb)	N = 18

Résultats

Sur 100 MMPI-2, 17 profils ont été considérés comme invalides. Ceux-ci concernent 11 patients avec un diagnostic de psychose, 3 un trouble de la personnalité, 1 un trouble des conduites alimentaires, 1 un trouble affectif et un dernier, un trouble mixte (axe I et axe II du DSM-III-R).

Avis des cliniciens sur la section « validité » du rapport

		N	*%*
A	Information insuffisante	11	13,3
B	Quelques informations	21	25,3
C	Information suffisante	30	36,1
D	Information plus que suffisante	15	18
E	Information approfondie	6	7,2

En résumé, l'avis des cliniciens sur la validité des 83 rapports qui leur sont parvenus a été dans l'ensemble favorable dans 61 % des cas, si l'on regroupe les informations jugées approfondies, plus que suffisantes et suffisantes.

Avis des cliniciens sur la section « tableau symptomatique » du rapport

		N	%
A	Information insuffisante	11	13,3
B	Quelques informations	23	27,7
C	Information suffisante	22	26,5
D	Information plus que suffisante	22	26,5
E	Information approfondie	5	6

En résumé, l'avis des cliniciens sur la qualité de la description des tableaux symptomatiques figurant dans les 83 rapports valides a été, dans l'ensemble, favorable dans 59 % des cas si l'on regroupe les informations jugées approfondies, plus que suffisantes et suffisantes.

Avis des cliniciens sur la section « relations interpersonnelles » du rapport

		N	%
A	Information insuffisante	10	12
B	Quelques informations	15	18
C	Information suffisante	26	31,3
D	Information plus que suffisante	22	26,5
E	Information approfondie	10	12

En résumé, l'avis des cliniciens sur la qualité de la section « relations interpersonnelles » du rapport figurant dans les 83 rapports valides a été, dans l'ensemble, favorable dans 70 % des cas si l'on regroupe les informations jugées approfondies, plus que suffisantes et suffisantes.

Avis des cliniciens sur la section « considérations diagnostiques » du rapport

		N	%
A	Information insuffisante	38	45,8
B	Quelques informations	14	16,9
C	Information suffisante	17	20,5
D	Information plus que suffisante	10	12
E	Information approfondie	4	4,8

En résumé, l'avis des cliniciens sur la qualité de la section « considérations diagnostiques » du rapport figurant dans les 83 rapports valides dans l'ensemble n'a été favorable que dans 37 % des cas si l'on regroupe les informations jugées approfondies, plus que suffisantes et suffisantes.

Avis des cliniciens sur la section « considérations thérapeutiques » du rapport

		N	%
A	Information insuffisante	14	16,9
B	Quelques informations	21	25,3
C	Information suffisante	27	32,5
D	Information plus que suffisante	16	19,3
E	Information approfondie	5	6

En résumé, l'avis des cliniciens sur la qualité de la section « considérations thérapeutiques » du rapport figurant dans les 83 rapports valides a été, dans l'ensemble, favorable dans 58 % des cas si l'on regroupe les informations jugées approfondies, plus que suffisantes et suffisantes.

Évaluation globale du rapport

À la question suivante : « Globalement, quel pourcentage des affirmations contenues dans le rapport vous apparaît comme pertinent pour décrire le patient? », les cliniciens ont répondu de la façon suivante :

		N	%
A	Moins de 20 %	10	12
B	De 20 à 39 %	17	20,5
C	De 40 à 59 %	18	21,7
D	De 60 à 79 %	18	21,7
E	De 80 à 100 %	20	24,1

La pertinence globale du rapport paraît suffisante pour 45,8 % des rapports si l'on retient comme satisfaisantes les catégories D et E, et pour 67,5 % si l'on est moins exigeant (catégories C, D et E).

Discussion

Ces divers résultats paraissent cohérents entre eux. Le nombre de rapports considérés comme parfaitement satisfaisants est d'environ 1 sur 4. À l'inverse, la fréquence des rapports dont la pertinence est estimée à moins de 40 % est de 32,5 %. À ces résultats d'ensemble, il faut ajouter que les diverses sections des rapports informatisés sont, pour les cliniciens, de qualité variable. La section pour laquelle il y a le plus de rapports considérés comme très insatisfaisants (catégorie A) est la section « considérations diagnostiques » (45,8 %). Puis suivent, d'assez loin les « considérations thérapeutiques » (16,9 %), le « tableau symptomatique » et la « validité » du rapport (13,3 %), enfin les « relations interpersonnelles » (12 %).

À l'inverse, la section pour laquelle il y a le plus de rapports considérés comme ayant fourni des informations très approfondies (catégorie E) est la section « relations interpersonnelles » (12 %), puis à un moindre degré, la section « validité » du rapport

(7,2 %), et « tableau symptomatique » (4,8 %). La comparaison du nombre de rapports très insuffisants (catégorie A) selon les grandes classes diagnostiques montre qu'aucun diagnostic n'est particulièrement sur-représenté (tableau 4).

Tableau 4. *Répartition des rapports insuffisants (catégorie A) selon les catégories diagnostiques*
(N = 38 sur 83 profils valides du MMPI-2 (35,8 %))

Diagnostic des cliniciens	*N patients*	*% sur 83*	*N rapports insuffisants*	*% sur 38*
Tr. dépressifs ou anxieux	18	21,7	9	23,7
Psychoses	19	22,9	10	26,3
Tr. isolé axe II (pers.)	22	26,5	5	13,2
Tr. mixte (axe I + axe II) ou 2 troubles axe I	21	25,3	12	31,6
Autres troubles	3	3,6	2	5,3
Total	83	100	38	100

Comparaison MMPI et MMPI-2

Une comparaison entre MMPI et MMPI-2 a pu être effectuée chez 39 patients. Chez 17 d'entre eux, le retest s'est situé à moins de trois mois de la première épreuve. Dans 21 cas sur 39, le profil du MMPI-2 était semblable à celui du MMPI. Dans 18 cas, le profil obtenu était nettement différent. Il est cependant impossible de se prononcer sur les similitudes ou les différences de l'état symptomatique des patients lors des deux cotations (absence d'utilisation d'une échelle d'évaluation psychopathologique au moment du test).

CONCLUSION

Cette première application du MMPI-2 à des patients français montre que la version française du questionnaire n'a pas soulevé de problèmes particuliers de compréhension de la part des malades. Le pourcentage de comptes-rendus ininterprétables en raison d'attitudes de réponses invalidant les résultats des échelles cliniques (17 %) correspond à ce que l'on observe habituellement dans la littérature pour ce type de questionnaires. L'avis des cliniciens sur la qualité du compte-rendu informatisé est plus favorable pour les sections « relations interpersonnelles », « tableau symptomatique » et « validité générale » que pour les sections « considérations thérapeutiques » et, surtout « considérations diagnostiques ». Ce type de résultat ne paraît guère surprenant puisque les réponses aux questions sont les seuls éléments pris en compte alors qu'un diagnostic en psychopathologie dépend aussi de l'histoire des troubles et d'une perspective obligatoirement diachronique. Ceci était d'ailleurs déjà clairement mentionné en 1986 par J. Perse dans le manuel en langue française du MMPI : « Le profil du MMPI ne doit pas être considéré comme étant essentiellement orienté vers le diagnostic psychiatrique... Le MMPI est sans doute moins un test de personnalité au sens strict du terme qu'un inventaire portant surtout sur des symptômes et des caractéristiques pathologiques. »

BIBLIOGRAPHIE

BERAH E., MIACH P., BUTCHER J., BOLZA J., COLAM S., MCASEY P. — *Computer-Based Interpretation of the MMPI-2.* An Australian Evaluation of the Minnesota Report. Paper given at the Australian Psychological Association Meetings, Melbourne, 1993.

BUTCHER J. N., PANCHERI P. — *Handbook of Cross-National MMPI Research.* Mn. University of Minnesota Press, Minneapolis, 1976.

BUTCHER J. N., GRAHAM J. R., WILLIAMS C. L., BEN-PORATH Y. S. — *Development and Use of the MMPI-2 Content Scales.* University of Minnesota Press, Minneapolis, 1990.

BUTCHER J. N. — *International Adaptations of the MMPI-2 : a Handbook of Research and Clinical Applications.* University of Minnesota Press, Minneapolis, in press.

DAHLSTROM W. G., WELSH G. S. — An MMPI Handbook. University of Minnesota Press, Minneapolis, 1960.

GILLET I., SIMON M., GUELFI J. D. — The MMPI-2 in France, Part I. In : J. N. BUTCHER (Ed). *International Adaptations of the MMPI-2.* University of Minnesota Press, Minneapolis, in press.

GUELFI J. D., BRUN A., MONIER C., SEUNEVEL F., SVARNA L. *et al.* — The MMPI-2 in France, Part II. In : J. N. BUTCHER (Ed). *International Adaptations of the MMPI-2.* University of Minnesota Press, Minneapolis, in press.

GRAHAM J. R. — *The MMPI : a Practical Guide (2nd ed).* Oxford University Press, Oxford, 1987.

GRAHAM J. R. — *MMPI-2 : Assessing Personality and Psychopathology.* Oxford University Press, New York, 1990.

HATHAWAY S. R., MCKINLEY J. C. — *MMPI. Minnesota Multiphasic Personality Inventory,* Manual, 1943, 1970. University of Minnesota, Minneapolis, Édition française : J. PERSE, Centre de psychologie appliquée, 1986 (1re édition, 1960).

HATHAWAY S. R., MCKINLEY J. C. — *MMPI-2. Minnesota Multiphasic Personality Inventory-2, Manual for Administration and Scoring.* University of Minnesota Press, Minneapolis, 1989.

MOREY L. S., WAUGH M. H., BLASHFIELD R. K. — MMPI Scales for DSM-III Personality Disorders : their Derivation and Correlates. *J. Pers. Assessment, 49,* 245-251, 1985.

WELSH G. S., DAHLSTROM W. G. (Eds) — *Basic Readings on the MMPI in Psychology and Medicine.* University of Minnesota Press, Minneapolis, 1956.

WIENER D. N. — Subtle and Obvious Keys for the MMPI. *J. Cons. Psychol., 12,* 164-170, 1948.

10

DU TEST D'INTELLIGENCE À LA NOUVELLE CLINIQUE

par R. D. DARDENNES et J. D. GUELFI

INTRODUCTION

L'évolution des idées sur les conceptions de l'intelligence depuis le début du siècle peut être illustrée par le chemin parcouru entre le premier test d'intelligence, celui de Binet et Simon, jusqu'aux épreuves inspirées par les modèles hiérarchiques et le cognitivisme.

Une évolution parallèle est observable dans le domaine de la psychopathologie. Notre objectif est de résumer les grandes lignes de l'évolution de la psychométrie de l'intelligence au cours du XXe siècle avant d'envisager les apports de la psychologie cognitive pour la clinique psychiatrique.

NAISSANCE ET PREMIERS DÉVELOPPEMENTS DE LA PSYCHOMÉTRIE DE L'INTELLIGENCE

La plupart des définitions modernes de l'intelligence se réfèrent soit à une « capacité générale d'adaptation à des situations nouvelles par des procédures cognitives » (M. Reuchlin), soit à l'aptitude à la pensée abstraite (J. Piaget). Les premiers tests mentaux comme ceux de J. McKeen Cattell ou de F. Galton cherchaient à explorer des processus élémentaires comme les sensations et la motricité plutôt que des processus intégratifs complexes. Binet et Simon, confrontés il y a 90 ans au problème de devoir sélectionner les enfants qui allaient s'avérer incapables de suivre une scolarité normale, ont voulu approcher l'intelligence de façon originale. Ils ont imaginé plusieurs séries de problèmes concrets auxquels les enfants peuvent être confrontés, susceptibles d'être résolus à partir de 3 ans pour les plus simples, comme le fait de donner son nom ou de montrer son nez, ou seulement vers 14-15 ans, comme le fait de distinguer des mots abstraits. Puis, ces diverses épreuves ont été classées par niveaux d'âge. Les réponses d'un enfant donné sont ensuite comparées aux résultats moyens obtenus par des groupes d'âges différents (groupes d'étalonnage). Par rapport aux groupes témoins, l'enfant examiné peut être en avance, en retard ou juste à son âge réel, chronologique. Le rapport de l'âge mental obtenu sur l'âge chronologique multiplié par 100 représente le quotient intellectuel. Il est normalement égal à 100. De nombreux tests d'intelligence furent construits sur ce modèle, comme, aux

États-Unis, le test de Terman, et, en France, la nouvelle échelle métrique de l'intelligence (NEMI) de Zazzo *et al.* en 1968.

Les mesures empiriques de Binet et de ses successeurs effectuées à l'aide d'épreuves très diverses sont sous-tendues par une représentation de l'intelligence conçue comme une aptitude générale évaluable à partir de la moyenne des résultats obtenus à des épreuves variées. La validité des tests d'intelligence de type de celui de Binet et Simon se juge sur la qualité du pronostic effectué et elle est, somme toute, assez rapide à vérifier. On peut globalement affirmer qu'à ce jour la valeur prédictive des tests d'intelligence — sans être négligeable — reste modérée (.30 à .50). En fait, la réussite scolaire est une donnée multifactorielle qui dépend de facteurs individuels comme l'intelligence, mais aussi de la motivation, des connaissances spécifiques, et de facteurs situationnels. Nombreux sont les spécialistes qui pensent que, malgré tout, la valeur prédictive des tests d'intelligence pourrait être améliorée par un meilleur échantillonnage des composantes de l'intelligence. D'autre part, les épreuves classiques évaluaient plus les résultats en termes de performances réalisées ou échouées plutôt que « les processus mis en œuvre pour les atteindre » et les tests « ne prennent généralement pas suffisamment en compte les connaissances déclaratives ou les stratégies. » (M. Huteau).

Les étapes déterminantes dans la voie du modèle factoriel de l'intelligence ont été illustrées par les travaux de C. Spearman, L. L. Thurstone, C. Burt, puis par le courant néo-piagétien, enfin par l'essor du cognitivisme.

Modèle factoriel, facteur g et modèle hiérarchique

C. Spearman a étudié les corrélations observées entre les résultats obtenus à différentes épreuves dites d'intelligence (épreuves de types différents) et a constaté que ces corrélations étaient susceptibles de s'expliquer par une source unique de variations interindividuelles : le facteur g. La voie du modèle factoriel de l'intelligence était ouverte. Tout se passe comme si « la réussite à une tâche donnée était conditionnée par une aptitude générale (commune à toutes les tâches) combinée à une aptitude spécifique à la tâche particulière. Pour chaque type de tâche, la contribution du facteur général… et celle du facteur spécifique varient. » (P. Pichot, 1962). L'appréciation de « ce facteur général » peut s'effectuer, schématiquement, de deux façons distinctes. Une première technique consiste à utiliser une épreuve unique que l'on sait être saturée en facteur général, par exemple une question pour laquelle la réponse exacte témoigne d'une bonne aptitude à former des concepts. La seconde technique consiste à utiliser conjointement des épreuves très variées en retenant la moyenne des notes, faisant alors l'hypothèse que les aptitudes spécifiques sont distribuées au hasard chez un individu et que la moyenne de toutes les épreuves témoigne surtout du facteur général.

Dans le modèle hiérarchique (C. Burt) dérivé de celui de Spearman, l'aptitude la plus générale est le facteur g ; puis viennent des facteurs de groupe — qui contribuent à un grand nombre d'activités — par exemple le facteur verbal et des facteurs spécifiques qui intéressent un nombre limité de tests. La réussite à un test déterminé est due en réalité à l'action simultanée de trois ordres de facteurs : général, de groupe et spécifique.

Aux États-Unis, L. L. Thurstone a été à l'origine d'une conception dite multifactorielle de l'intelligence. La réussite à une épreuve dépend de la conjugaison de l'action

de plusieurs facteurs de rang équivalent. Or, les facteurs ne sont pas indépendants les uns des autres. Il est possible d'effectuer une deuxième analyse factorielle portant cette fois sur les corrélations entre facteurs. Les corrélations existantes entre les différentes aptitudes mentales primaires sont sous la dépendance d'un facteur général dit de second ordre. Ainsi, la batterie d'aptitudes mentales primaires de Thurstone (1941) comportait 7 facteurs :

S : facteur spatial (habileté à visualiser des relations spatiales).

P : facteur perceptif (identification d'un détail noyé dans un complexe).

N : facteur numérique (facilité pour le calcul).

V : facteur verbal (signification des mots).

W : facteur verbal (aptitude à utiliser des mots dans un cadre formel).

M : facteur mnémonique.

R : facteur de raisonnement.

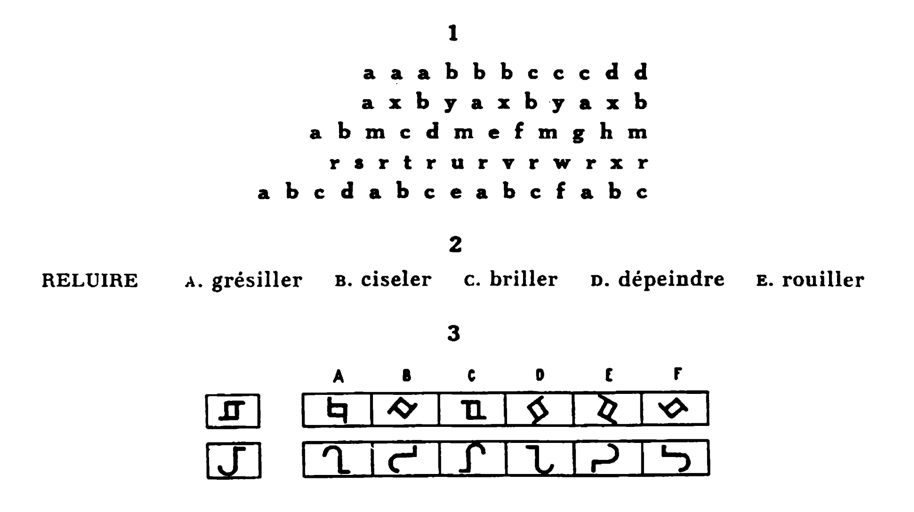


Fig. 1. *Trois exemples de tests factoriels d'aptitudes*, empruntés à la batterie d'aptitudes mentales primaires de Thurstone ou PMA (avec l'autorisation du Centre de psychologie appliquée).

Les trois exemples de la figure 1 correspondent à l'exploration de 3 facteurs saturés en facteur général de second ordre :

– Raisonnement : trouver la lettre qui continue chaque série.

– Verbal : désigner parmi A, B, C, D ou E, le synonyme du mot de gauche.

– Spatial : désigner parmi les 6 figures de droite celles décrites dans la figure de gauche par rotation dans le plan et celles qui sont la figure symétrique en miroir.

Plusieurs tests construits par M. Reuchlin et E. Valin en 1958, à savoir quinze épreuves inspirées de 4 des aptitudes primaires mentionnées ci-dessus, permettent de retrouver quatre facteurs de groupe et un facteur général rendant compte de 50 à 77 % de la variance.

Un autre test, saturé en facteur général mais dans lequel intervient aussi un facteur de groupe Spatial, est le *Progressive Matrices de Raven PM-38* (1938).

Ce test présente des séries de 4 ou 9 dessins géométriques dont manque le dernier élément. La tâche consiste à découvrir la loi qui définit les rapports entre les différents éléments du tableau et à désigner l'élément manquant parmi différents choix proposés sous le tableau (fig. 2). Soixante tableaux sont ainsi présentés (5 séries de 12, de difficulté croissante). Le temps est libre ou limité à 20 minutes. Il existe un étalonnage en centilage de six ans à l'âge adulte.

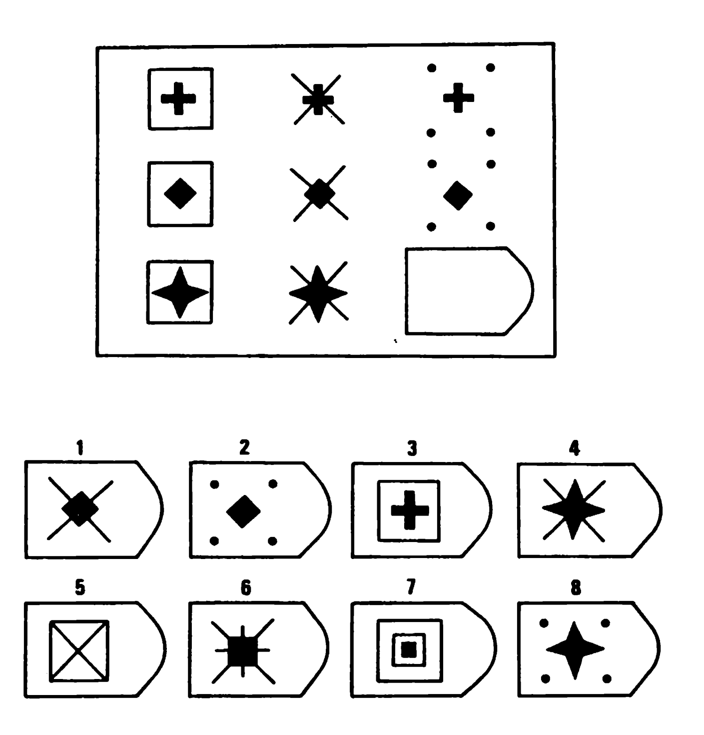
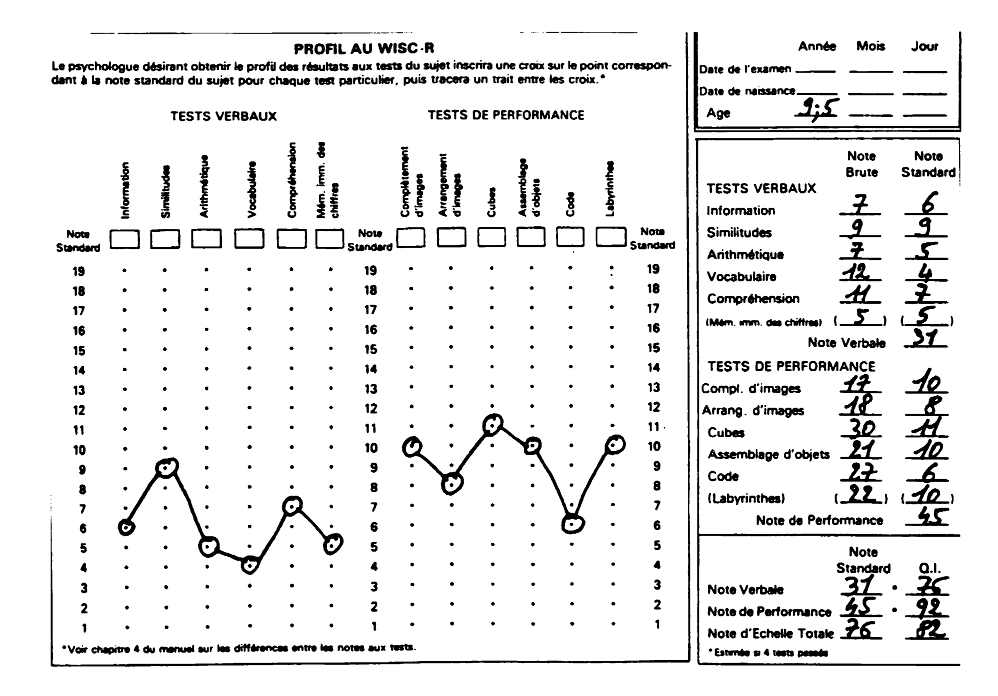


Fig. 2. *Exemple de tableau du Progressive Matrices de Raven:* Sanglade-Andronikof A. et Verdier-Gibello M.-L. – L'examen psychologique de l'enfant. Les tests d'intelligence, d'aptitude, de raisonnement. *Encycl. Méd. Chir., Paris, Psychiatrie, 37180,* C 10, 2-1983. (Reproduit avec l'autorisation de l'EMC.)

Le modèle hiérarchique a été retenu par de nombreux auteurs comme P. E. Verron ou R. B. Cattell. Sa pertinence peut s'illustrer par l'analyse des données d'un test composite comme celui de Wechsler. Il en existe trois versions : une forme adaptée à la période préscolaire et primaire, *Wechsler Preschool and Primary Scale of Intelligence* ou WPPSI pour enfants de 4 à 6-7 ans, une forme pour enfants (de 6 ans 1/2 à 16 ans 1/2), révisée en 1974, traduite en français en 1981, *Wechsler Intelligence Scale for Children-Revised* ou WISC-R, et la forme pour adultes *Wechsler Adult Intelligence Scale* ou WAIS. Toutes les analyses effectuées sur les 12 sub-tests parviennent à distinguer deux facteurs de groupe : le *facteur verbal* et le *facteur performance* (fig. 3).



PROFIL AU WISC-R

Le psychologue désirant obtenir le profil des résultats aux tests du sujet inscrira une croix sur le point correspondant à la note standard du sujet pour chaque test particulier, puis tracera un trait entre les croix.*

TESTS VERBAUX — TESTS DE PERFORMANCE

Information, Similitudes, Arithmétique, Vocabulaire, Compréhension, Mém. imm. des chiffres — Note Standard — Complètement d'images, Arrangement d'images, Cubes, Assemblage d'objets, Code, Labyrinthes — Note Standard

Note Standard: 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1

*Voir chapitre 4 du manuel sur les différences entre les notes aux tests.

	Année	Mois	Jour
Date de l'examen			
Date de naissance			
Age	9;5		

	Note Brute	Note Standard
TESTS VERBAUX		
Information	7	6
Similitudes	9	9
Arithmétique	7	5
Vocabulaire	12	4
Compréhension	11	7
(Mém. imm. des chiffres)	(5)	(5)
Note Verbale		31
TESTS DE PERFORMANCE		
Compl. d'images	17	10
Arrang. d'images	18	8
Cubes	30	11
Assemblage d'objets	21	10
Code	27	6
(Labyrinthes)	(22)	(10)
Note de Performance		45

	Note Standard	Q.I.
Note Verbale	31	76
Note de Performance	45	92
Note d'Echelle Totale	76	82

*Estimée si 4 tests passés

Fig. 3. *Exemple de profil obtenu au WISC-R :* Sanglade-Andronikof A. et Verdier-Gibello M.-L. – L'examen psychologique de l'enfant. Les tests d'intelligence, d'aptitude, de raisonnement. *Encycl. Méd. Chir., Paris, Psychiatrie, 37180,* C 10, 2-1983. (Reproduit avec l'autorisation de l'EMC.)

LE MODÈLE UNIDIMENSIONNEL DE J. PIAGET

Le modèle de J. Piaget ne porte pas, lui, sur les différences inter-individuelles et il est, dans son essence, unidimensionnel. Le développement de l'intelligence se caractérise par « une marche vers l'équilibre » au cours de laquelle tous les enfants franchissent une série de stades se présentant pour tous dans le même ordre et « les enfants ne diffèrent les uns des autres que par la vitesse avec laquelle ils progressent sur une voie unique » (M. Reuchlin et F. Bacher, 1989). La méthode d'exploration opératoire de J. Piaget repose sur des tâches dépendant peu des acquisitions culturelles ou scolaires. Les principales épreuves explorent les capacités : de *conservation* (tests où l'on doit découvrir des invariants dans les qualités ou les propriétés physiques d'objets), de représentation de l'espace, de perception du temps, de la qualité des images mentales (représentation de configurations, de mouvements et de transformations), des opérations logiques (étude des rapports entre objets : classification, sériations, nombres), du langage enfin.

Des instruments d'évaluation psychométrique inspirés par la théorie de J. Piaget ont été développés à Genève par B. Inhelder, en France par L. Longeot et par l'équipe franco-suisse animée par F. Rieben, A. de Ribaupierre et J. Lautrey.

Les corrélations trouvées entre les différentes épreuves piagétiennes et les QI de Weschler sont élevées : .80 pour le QI verbal, .82 pour le QI performance (L. G. Humphreys, 1980). Cependant, l'hypothèse d'un développement unidimensionnel

de l'intelligence « se trouve en contradiction avec certains autres faits » pour F. Bacher. Et c'est aussi à ce type de résultats que sont parvenus nombre de continuateurs de l'œuvre de J. Piaget comme M. Laurendeau et A. Pinard, F. Longeot ou J. Lautrey *et al.* Pour ce dernier auteur, par exemple, qui va nous permettre d'aborder le cognitivisme, les processus mis en jeu dans le développement peuvent prendre des importances relatives variées chez différents individus. Il distingue deux grands types de traitement de l'information : un *mode analogique* (le plus efficace dans les domaines infralogiques) et un *mode propositionnel* (le plus efficace dans le domaine logico-mathématique). Certains apprentissages qui favorisent les interactions entre les deux processus semblent favorables au développement cognitif.

LES NÉO-PIAGÉTIENS ET LE COGNITIVISME

Divers chercheurs néo-piagétiens ont intégré à la théorie de Piaget certains concepts fonctionnalistes du traitement de l'information comme ceux de la mémoire de travail ou des stratégies. Le point de vue cognitiviste est en effet souvent centré sur les modalités du traitement de l'information. Pour E. Hurt, par exemple, la réussite à une épreuve complexe de compréhension dépend du bon fonctionnement de divers modules élémentaires ; chacun a une fonction précise : capacité de la mémoire à court terme, modalités de conservation d'un agencement déterminé des informations, vitesse d'accès dans la mémoire à long terme.

D'autre part, les stratégies utilisées pour le fonctionnement des modules élémentaires s'avèrent fort variables d'un individu à l'autre. Les points de vue des « processus directeurs » et des « métacomposantes », centraux dans la théorie de R. J. Sternberg dite théorie componentielle, permettent d'illustrer ce propos. L'activité mentale est décrite par cet auteur à l'aide de trois types d'opérations.

Une composante est un processus élémentaire de traitement de l'information opérant sur des représentations internes ou des symboles. R. J. Sternberg distingue :

1) les métacomposantes ou processus « directeurs » (executive) utilisés pour planifier le déroulement d'une tâche, par exemple la sélection et le déclenchement d'une stratégie de combinaison optimale de composantes inférieures pour résoudre un problème, la distribution optimale de « l'attention », etc. ;

2) les composantes de performance, chargées de l'exécution d'une tâche ;

3) les composantes d'acquisition des connaissances enfin, dont trois catégories correspondent à trois phases des apprentissages : l'encodage sélectif, les connaissances sélectives et l'intégration de ces ensembles aux connaissances antérieures.

Ce système hiérarchique a une capacité limitée. Un problème nouveau mobilise l'ensemble du système. Un problème familier voit se transférer l'activité du système global conscient à des systèmes « locaux » fonctionnant de façon automatique. Les sujets se distinguent non seulement par leurs aptitudes variables à traiter des problèmes nouveaux mais aussi par leur aptitude à automatiser le traitement de l'information. La différence principale entre sujets d'intelligence normale et retardée se situerait plus au niveau d'une métacomposante du processus de direction que des processus élémentaires.

Existe-t-il aussi des différences dans la vitesse du traitement de l'information? Des recherches dites corrélationnelles ont étudié conjointement la réussite à des épreuves cognitives et divers paramètres comme le temps de réaction ou la vitesse d'exploration de la mémoire à court terme ou encore la vitesse d'accès à la mémoire

à long terme (comme dans le paradigme de M. Posner). Lorsqu'un sujet doit décider de la similitude physique (AA) ou sémantique (Aa) de deux lettres, la différence de latence de sa réponse dans ces deux situations peut être considérée comme un reflet de « la rapidité d'accès aux codes sémantiques ». Cependant, ces travaux intéressants ont leurs limites. On vient de voir, par exemple, qu'un temps de latence dans l'exécution d'une tâche était identifié au temps du fonctionnement d'un processus hypothétique. Or, rien ne permet actuellement de vérifier le bien-fondé de cette hypothèse. Enfin, les corrélations entre test d'intelligence et capacités élémentaires restent faibles, de l'ordre de .30.

SYNTHÈSE PROVISOIRE

Les modèles factoriels comportant un facteur général et des facteurs de groupe éventuellement hiérarchisés en plusieurs niveaux sont ceux qui s'avèrent les plus pertinents pour rendre compte des corrélations entre différentes épreuves d'intelligence. La nécessité de faire intervenir des « processus directeurs », des métacomposantes, définissant et contrôlant à un niveau supérieur les stratégies employées dans la mise en œuvre de processus plus élémentaires s'est progressivement imposée. Cependant, il est deux conceptions bien différentes du modèle hiérarchique en matière de développement cognitif. Dans une première conception, les facteurs identifiés sont indépendants les uns des autres ; le facteur g, par exemple, est indépendant des facteurs de groupe.

Dans l'autre, il existe des relations fonctionnelles de dépendance ou d'interdépendance entre les niveaux et les processus hiérarchiquement supérieurs « exercent une influence, un contrôle sur les processus situés au-dessous ». D'autre part, « une même note en facteur général peut découler de réussites différentes dans les facteurs de groupe. Des composantes ou processus différents peuvent être mis en œuvre par des stratégies différentes permettant toutes de résoudre le problème posé. » (F. Bacher).

Enfin, les performances ne dépendent pas que de paramètres structuraux. Elles dépendent aussi de la *quantité « d'attention »* disponible et les facteurs qui déterminent l'importance de cette quantité d'énergie ou d'attention restent à préciser.

Il existe enfin d'autres points de vue qui mériteraient d'être développés comme la théorie d'inspiration neuropsychologique de Das et Naglieri inspirée de Luria avec trois facteurs à la base de l'activité intellectuelle : *l'attention* qui permet le maintien d'une vigilance adéquate, un second système étant responsable du *codage de l'information*, le 3e dans la région frontale, qui *planifie* et *régule* l'exécution des tâches et le contrôle de l'impulsivité.

Diverses épreuves peuvent être proposées qui correspondent à ces différents facteurs, par exemple : pour la planification, des épreuves faisant appel aux capacités d'anticipation ; pour l'attention, l'appariement des lettres ; pour le codage simultané : une épreuve de compréhension verbale comme les progressive matrices de Raven et pour le codage successif : une épreuve de compréhension de syntaxe.

ÉVALUATION DE L'INTELLIGENCE ET DE LA PSYCHOPATHOLOGIE : UNE ÉVOLUTION PARALLÈLE

L'évaluation de l'intelligence et celle de la psychopathologie ont suivi une évolution parallèle à deux titres.

D'une part, ces deux domaines ont été transformés par l'approche cognitive. Cette perspective a orienté la clinique vers des éléments susceptibles de rendre compte du traitement de l'information opéré par les patients souffrant d'un trouble psychique. Elle a suscité une analyse différente des éléments sémiologiques classiques et une recherche des caractéristiques de traitement de l'information susceptibles de décrire ces troubles.

D'autre part, ce point de vue cognitif a contribué à modifier sensiblement les buts de l'évaluation des troubles psychiques. Comme pour l'évaluation de l'intelligence, il s'opère un glissement de finalité de l'évaluation d'un but diagnostique à un but de remédiation (Huteau, 1990). Il ne s'agit plus seulement de classer le patient dans une catégorie diagnostique, catégorie qui résume les informations nécessaires à sa prise en charge et à son pronostic, mais aussi de repérer cliniquement les caractéristiques cognitives dont la modification permet de traiter le trouble lui-même ou les handicaps qui en résultent.

Analyse cognitive des symptômes psychopathologiques

Hallucinations et délire

Ces deux types de symptômes fréquents mais non spécifiques — on peut les observer tant dans les troubles schizophréniques que dans les troubles affectifs, sans parler des délires chroniques et des psychoses hallucinatoires chroniques — ont fait l'objet d'études empiriques sur les mécanismes cognitifs qui pourraient les sous-tendre.

Les hallucinations sont-elles dues à une incapacité à différencier les événements réels des événements imaginaires ?

C'est l'hypothèse que fait Bentall (1990) lors de sa revue des théories des hallucinations et des expériences menées dans ce domaine. Il postule une défaillance d'ordre métacognitif, à savoir l'incapacité de discriminer les événements mentaux des événements externes (*reality discrimination* ou *reality testing*). Une telle méprise entre réalité et imagination peut être provoquée expérimentalement chez des sujets sains. Chez les patients hallucinés, plusieurs observations étayent cette hypothèse : dans les épreuves de détection de signal, ils ont un taux élevé de faux-positifs (biais de détection) ; quand on leur demande de distinguer deux signaux, ils ont une rapidité et une confiance excessives dans leurs jugements. À capacités mnésiques égales, patients hallucinés et non hallucinés diffèrent dans l'attribution de leurs souvenirs. Alfred Heilbrun (1980) a montré que les patients ayant de telles hallucinations sont moins capables de reconnaître les pensées qu'ils ont personnellement exprimées lors d'un entretien précédent que des patients sans hallucinations. Cet effet est observé alors que les capacités mnésiques et la stabilité des opinions émises sont semblables pour les deux groupes (lors de l'entretien on demandait au patient d'exprimer ses opinions sur l'amitié, l'honnêteté des hommes politiques, la religion, l'inégalité des

sexes, et l'argent). Cette difficulté à reconnaître ses propres pensées serait relativement spécifique des sujets hallucinés; dans une expérience analogue, Bentall *et al.* (1991) montrent que les schizophrènes hallucinés diffèrent des délirants non hallucinés dans leurs erreurs : ils attribuent plus volontiers leurs propres réponses à l'expérimentateur. Une explication possible à cette difficulté à distinguer réalité externe et pensées est une mauvaise utilisation d'indices utiles à cette distinction. Cette explication est suggérée par l'expérience d'Heilbrun (1983) où les sujets hallucinés auditifs avaient des difficultés à localiser la direction des sons.

Le délire : rationalisation d'expériences aberrantes ou erreurs de raisonnement?

Les convictions délirantes sont-elles, comme le suggère Maher, le produit d'un raisonnement normal appliqué à des expériences perceptuelles aberrantes? Et/ou le raisonnement des sujets délirants est-il erroné? Sur la base du modèle cognitif de la dépression de Beck, les convictions délirantes ont fait l'objet d'analyses analogues à celles utilisées pour mettre en évidence les schémas de pensée des déprimés. Bentall et son équipe ont ainsi montré des résultats similaires à ceux observés chez les déprimés, à savoir un meilleur rappel des informations menaçantes (des histoires de personnes suivies ou observées) et un biais attentionnel envers les informations à caractère persécutif lors d'un test de Stroop modifié (dans le Stroop modifié, les mots colorés ne sont pas des noms de couleur mais des mots choisis pour leur forte connotation anxieuse, dépressive ou, comme ici, persécutoire; ex. «espion», «suivre»). D'autre part, les sujets délirants ont une très nette tendance à faire des conclusions rapides. À titre d'exemple, Garety *et al.* (1991) comparent des sujets délirants (schizophrènes paranoïdes et paranoïaques) et non délirants (normaux ou ayant un trouble anxieux) sur une tache de décision. Il y a des urnes contenant des perles de deux couleurs différentes dans une proportion connue. On leur présente successivement des perles dans une séquence donnée jusqu'à ce qu'ils décident de quelle urne proviennent en majorité ces perles. Les sujets délirants se décident en majorité dès la première perle, tandis qu'il faut une série de 3 à 5 perles pour que les sujets non délirants se prononcent sur leur urne d'origine. Par contre, leur conviction, que l'on évalue par la confiance en leur décision, n'est pas inébranlable et varie dès qu'on leur présente un indice contradictoire (c'est-à-dire une perle dont la couleur ne correspond pas à leur hypothèse sur l'urne) tandis qu'elle est plus stable chez les sujets non délirants. Ce type d'approche tente donc de décomposer les paramètres des convictions délirantes et le type de «distorsion cognitive», pour reprendre la terminologie de Beck, mis en jeu. Cependant, ces travaux n'apportent pas d'éclairage sur le contenu des hallucinations et des délires. Et il reste à savoir si une intervention centrée sur ces phénomènes peut amener une réduction de ces symptômes.

Remédiation cognitive

La thérapie cognitive de la dépression a élargi le champ de la clinique de la dépression. Les pensées du déprimé, sa représentation de lui-même et de ses interactions avec l'environnement et ses jugements sont devenus des éléments courants de l'évaluation d'un état dépressif. Cet élargissement de la clinique de la dépression a également permis un abord psychothérapeutique en apprenant au patient à faire face à

ses idées négatives. Cette approche s'est généralisée à de nombreuses pathologies psychiatriques et elle a surtout renouvelé l'abord de l'anxiété chronique.

Inquiétudes permanentes : voie d'abord de l'anxiété chronique

L'approche de la classique névrose d'angoisse a été transformée par l'examen des contenus de pensée des patients ayant une anxiété dite « flottante ». D'abord décrites par Beck en 1974, ces pensées anxieuses ont suscité un grand intérêt et sont plus connues sous le terme de *Worry* que l'on peut traduire par *inquiétudes*. Ces inquiétudes excessives et/ou irréalistes sont devenues la caractéristique fondamentale de l'anxiété généralisée définie dans le DMS-III-R.

L'équipe de Borkovec a publié de nombreuses expérimentations sur ce phénomène des inquiétudes (une synthèse de leurs travaux est publiée dans Rapee, 1991, chapitre 2) : ce sont principalement des verbalisations incontrôlables tentant de résoudre une éventualité malheureuse telle que la maladie d'un proche. Cette équipe a constaté que les sujets inquiets avaient moins d'images mentales que des témoins mais qu'ils en avaient autant après traitement. D'autre part, si l'on demande à des sujets ayant une anxiété de performance de s'inquiéter à propos d'une éventuelle prise de parole en public, avant d'imaginer visuellement la scène, leur réaction cardiaque est nulle; alors que la fréquence cardiaque augmente si la période d'imagination est précédée d'une phase de détente. Ces deux observations ont conduit ces auteurs à concevoir l'inquiétude comme un moyen d'inhiber le traitement de l'information émotionnelle, moyen qui a également pour conséquence de maintenir cette inquiétude. Ce phénomène est devenu une cible importante des thérapies cognitives de l'anxiété généralisée où l'on opère une exposition aux inquiétudes (Barlow, 1993).

Tentatives de remédiation neuropsychologique dans la schizophrénie : apprendre le WISC

Ce dernier exemple illustre un tournant dans la finalité de l'approche cognitive de la schizophrénie. Les altérations cognitives des schizophrènes sont connues depuis longtemps mais l'idée d'agir directement sur les troubles de l'attention, de la mémoire, etc., pour diminuer les symptômes et les handicaps du schizophrène est une idée récente et encore controversée (Green, 1993). Une partie du débat tourne autour de la perte de « souplesse cognitive » mise en évidence chez le schizophrène avec le test WISC.

Ce test consiste à trier des cartes selon une des trois catégories suivantes : la couleur, la forme ou le nombre des symboles géométriques qui figurent sur celles-ci. Après 6 ou 10 bonnes réponses selon une des stratégies (couleur par ex.), l'expérimentateur change la règle du jeu sans prévenir le sujet (par ex. passage à une règle d'appariement sur la forme et non plus sur la couleur des symboles). On mesure le nombre de règles trouvées et le nombre de persévérations, c'est-à-dire de persistance à trier les cartes selon une règle dont l'expérimentateur a clairement informé le sujet qu'elle n'est plus en vigueur. Ces persévérations sont caractéristiques d'une altération frontale dorsolatérale et s'observent également de façon assez nette dans la schizophrénie. On imagine que ce manque de souplesse puisse entraîner un handicap sur le plan du comportement. Parmi les tentatives de remédiation cognitive, celle publiée récemment par Ann Delahunty *et al.* (1993) illustre bien cette démarche. Trois sujets ont été soumis à un entraînement de 15 jours sur plusieurs tâches

(poursuite oculaire, entraînement à distinguer des illusions visuelles, développement de la motricité fine et jeux analogues au WISC avec renforcement par jetons). Les résultats ont montré une normalisation durable (6 mois) des performances au WISC. Cette « rééducation » n'a pas eu d'effets sur la symptomatologie générale, ni sur les capacités sociales du sujet, mais elle est exemplaire d'un courant partisan de traiter les déficits cognitifs, à la manière des neuropsychologues.

CONCLUSION

La psychopathologie clinique, tout comme la conception de l'intelligence, a été transformée par le paradigme cognitif. Il faut cependant noter que cette approche reste encore grossière au niveau de la clinique psychiatrique si l'on se réfère au modèle de traitement de l'information qui reste très schématique, pour ne pas dire métaphorique, dans les travaux présentés. Cependant, l'approche cognitive par sa description transversale des mécanismes du fonctionnement psychique offre des possibilités de remédier aux altérations des éléments de la chaîne de transformation de l'information, que ce soit dans le domaine psychiatrique ou dans le domaine des déficits intellectuels. En ce sens, elle modifie sensiblement les buts de l'évaluation dans ces deux domaines : il ne s'agit plus seulement de classer un individu, mais aussi de l'aider à remédier à ses éventuels dysfonctionnements. C'est dans ce sens que seront amenés à se développer les futurs outils d'évaluation dans ces deux domaines.

BIBLIOGRAPHIE

BARLOW D. H. — *Clinical Handbook of Psychological Disorders* (2nd éd.), The Guilford Press, New York, 1993 (chap. 4 « Anxiété généralisée » et chap. 6 « Dépression »).

BECK A. T. — Ideational Components of Anxiety Neurosis. *Archives of General Psychiatry, 31*, 319-325, 1974.

BENTALL R. P. — The Illusion of Reality : a Review and Integration of Psychological Research on Hallucinations. *Psychological Bulletin, 107*, 82-95, 1990.

BUTLER R. W., BRAFF D. L. — Delusions : a Review and Integration. *Schizophrenia Bulletin, 17*, 633-647, 1991.

DELAHUNTY A., MORICE R., FROST B. — Specific Cognitive Rehabilitation in Schizophrenia. *Psychological Medicine, 23*, 221-227, 1993.

DELAY J., PICHOT P. — *Abrégé de psychologie à l'usage de l'étudiant*. Masson, Paris, 1962.

EYSENCK M. W. — *Anxiety. - The Cognitive Perspective*. Lawrence Erlbaum Associates, Hillsdale, 1992.

GARETY P. A., HEMSLEY D. R., WESSELY S. — Reasoning in Deluded Schizophrenics and Paranoid Patients. Biases in Performance on a Probabilistic Inference Task. *The Journal of Nervous and Mental Disease, 179*, 194-201, 1991.

GIBELLO B. — *L'enfant à l'intelligence troublée* (2nd éd.). Éd. du Centurion, 1984.

GOTLIB I. H., HAMMEN C. L. — *Psychological Aspects of Depression. Toward a Cognitive-Interpersonal Integration*. John Wiley & Sons, Chichester, 1992.

GREEN M. F. — Cognitive Remediation in Schizophrenia : Is It Time Yet? *Am. J. of Psychiatry, 150*, 178-187, 1993.

HEILBRUN A. B. — Impaired Recognition of Self-Expressed Thoughts in Patients with Auditory Hallucinations. *Journal of Abnormal Psychology*, *89*, 728-736, 1980.

HEILBRUN A. B., BLUM N., HAAS M. — Cognitive Vulnerability to Auditory Hallucination. Preferred Imagery Mode and Spatial Location of Sounds. *Br. J. of Psychiatry*, *143*, 294-299, 1983.

HUMPHREYS L. G., RICH S. A., DAVEY T. C. — A Piagetian Test of General Intelligence. *Development Psychology*, *22*, 243-677, 1987.

HUTEAU M. — La variété des formes et des composantes de l'intelligence. In : M. REUCHLIN, J. LAUTREY, C. MARANDAZ, T. OHLMANN (Eds), *Cognition : l'individuel et l'universel*, PUF, Paris, 253-283, 1990.

LAUTREY C., HUTEAU M. — L'évaluation du développement et des compétences cognitives chez l'enfant. Quoi de neuf? *Rev. Psychol. Appl.*, *40*, 437-464, 1990.

LONGEOT F. — *Psychologie différentielle et théorie opératoire de l'intelligence*. Dunod, Paris, 1969.

LONGEOT F. — *Les stades opératoires de Piaget et les facteurs d'intelligence*. Presses Universitaires de Grenoble, Grenoble, 1978.

LONGEOT F. — *L'échelle de développement de la pensée logique* (EPL). Éditions Scientifiques et Psychologiques, Issy-les-Moulineaux, 1979.

MATHEWS A. — Why Worry? The Cognitive Function of Anxiety. *Behaviour Research and Therapy*, *28*, 455-468, 1990.

NAGLIERI J. A., DAS J. P. — A Cognitive Processing Theory of the Measurement of Intelligence. *Educational Psychologist* (cité par J. LAUTREY C. et M. HUTEAU, 1990), *24*, 185-206, 1989.

PIAGET J. — *La naissance de l'intelligence chez l'enfant*, Delachaux et Niestlé, Neuchâtel, 1936.

PICHOT P. — *Les tests mentaux*. Que Sais-Je? n° 626, PUF, Paris, 1re édition : 1954.

RAPEE R. M., BARLOW D. H. — Chronic Anxiety. – *Generalized Anxiety Disorder and Mixed Anxiety-Depression*. The Guilford Press, New York, 1991.

REUCHLIN M., BACHER F. — *Les différences individuelles dans le développement cognitif de l'enfant*. PUF, Paris, 1989.

RIEBEN L., DE RIBAUPIERRE A., LAUTREY J. — *Le développement opératoire de l'enfant entre 6 et 12 ans*. Paris, CNRS, 1983.

SANGLADE-ANDEONIKOF A., VERDIER-GIBELLO M. L. — *L'examen psychologique de l'enfant. Les tests d'intelligence, d'aptitude, de raisonnement*. Encycl. Médico-Chir. (Paris) *Psychiatrie*, *37*, 180 C 10, 2, 1983.

STERNBERG R. J. — *Intelligence, Information Processing, and Analogical Reasoning : the Componential Analysis of Human Ability*. L. Erlbaum, Hillsdale, 1977.

STERNBERG R. J. — Beyond IQ, a Triarchic Theory for the Measurement of Human Intelligence. Freeman, New York, 164-185, 1984.

STERNBERG R. J. — Mechanisms of Cognitive Development : a Componential Approach. In : R. J. STERNBERG (Ed.), Mechanisms of Cognitive Development, Freeman, New York, 1984.

STRAUSS M. E. — Relations of Symptoms to Cognitive Deficits in Schizophrenia. *Schizophrenia Bulletin*, *19*, 215-231, 1993.

WECHSLER D. — *La mesure de l'intelligence de l'adulte* (1944). PUF, Paris, 1956.

WECHSLER D. — *Échelle d'intelligence de Wechsler pour enfants*. Forme révisée, CPA, Paris, 1981.

11

TROUBLE ANXIEUX ET DÉPRESSIF MIXTE : UNE NOUVELLE ENTITÉ

par J.-P. LÉPINE

Au plan clinique, l'intrication entre manifestations anxieuses et dépressives est habituelle et les diverses théories et classifications des troubles psychopathologiques se sont attachées au fil des époques ou des diverses écoles psychiatriques tantôt à vouloir les différencier ou, à l'inverse, à les envisager dans une perspective plus unitaire selon un gradient de sévérité. Les tendances actuelles, reflétées dans les classifications internationales récentes, ont principalement mis l'accent, dans une approche catégorielle, sur les distinctions qui pouvaient les opposer selon des schémas diagnostiques hiérarchisés. Cependant, il est rapidement apparu que les divers troubles tels qu'ils sont décrits par exemple dans la classification du DSM-III ne pouvaient être considérés uniquement comme des catégories exclusives (ou disjointes) et également que la symptomatologie de nombreux patients ne satisfaisait pas aux descriptions cliniques proposées, alors même qu'un handicap avéré ou une souffrance psychologique certaine les amenait à consulter pour leurs troubles. Ces deux éléments qui, pour certains, auraient une importance encore plus grande en médecine générale qu'en psychiatrie (Barrett *et al.*, 1988 ; Katon et Roy Byrne, 1991) justifieraient l'inadéquation supposée des classifications actuelles et rendraient nécessaires des classifications adaptées à la médecine générale, cadre dans lequel les troubles psychopathologiques seraient fort différents de ceux observés en psychiatrie. Des vocables divers sont d'ailleurs largement utilisés et notamment celui de troubles anxio-dépressifs dans des acceptions souvent variables d'un clinicien à l'autre en se référant à des modèles théoriques également multiples. Les approches thérapeutiques actuelles qu'elles soient chimiothérapiques ou cognitivo-comportementales mettent également l'accent dans de nombreux cas sur une similitude d'efficacité des mêmes stratégies dans les troubles anxieux ou dépressifs et ne contribuent que fort peu à une clarification de ce problème.

Nous envisagerons dans une perspective clinique et théorique, les principaux points permettant de mieux préciser des notions ou des concepts couramment utilisés ainsi que les arguments pouvant étayer les propositions de définition du cadre nouveau du trouble anxieux et dépressif mixte tel qu'il a été proposé récemment dans les dernières révisions des classifications internationales ainsi que des syndromes cliniques qui peuvent s'en rapprocher.

RELATIONS ANXIÉTÉ-DÉPRESSION

L'association entre anxiété et dépression a été conceptualisée dans de nombreuses perspectives. Selon Clark (1989), il pourrait s'agir d'éléments cliniques différents évoluant selon un même continuum de sévérité, de manifestations alternatives d'une diathèse commune sous-jacente, de syndromes hétérogènes qui seraient associés en raison de l'existence de sous-types communs à ces deux syndromes, de phénomènes distincts pouvant évoluer de l'un à l'autre dans le temps mais également de phénomènes distincts au plan conceptuel et empirique. Divers travaux étayent l'une ou l'autre de ces hypothèses.

Cette association entre symptômes et troubles anxieux et dépressifs renvoie aux concepts de *co-occurrence* et de *comorbidité*. Selon Feinstein (1970), un trouble est comorbide avec un autre trouble dès lors qu'il correspond « à toute entité clinique distincte et additionnelle qui a existé ou qui peut survenir durant l'évolution clinique d'un patient présentant la maladie index étudiée. » Cette définition implique une relation variable dans le temps d'éléments cliniques envisagés au plan de la maladie, des troubles ou à tout le moins de syndromes qui peuvent coexister ou se succéder l'un l'autre. On parlera plus volontiers de co-occurrence symptomatique lorsque, à un instant donné, un patient présente des symptômes du registre des deux troubles.

CO-OCCURRENCE SYMPTOMATIQUE

L'association des symptômes anxieux et dépressifs est extrêmement fréquente. La plupart des échelles dimensionnelles évaluant l'anxiété et la dépression sont hautement corrélées entre elles, ce qui suggère une absence de validité discriminante entre elles. Si l'on envisage, à titre d'exemple, les échelles d'Hamilton anxiété et dépression de nombreux items tels les troubles du sommeil, l'anxiété, l'anxiété somatique, les symptômes gastro-intestinaux, les symptômes somatiques généraux et l'humeur dépressive sont communs à ces deux échelles.

De plus, au plan catégoriel, dans les nouvelles classifications internationales, comme le DSM-III-R, certains symptômes tels la fatigue ou fatigabilité, la fébrilité, les difficultés de concentration, l'irritabilité, les troubles du sommeil et de nombreux symptômes somatiques sont retenus en tant que critères diagnostiques à la fois pour le diagnostic de l'anxiété généralisée et de l'épisode dépressif.

Ce chevauchement symptomatique entre anxiété et dépression peut rendre compte des phénomènes de co-occurrence rapportés dans de nombreuses études. Il peut également correspondre à l'existence d'un noyau commun de souffrance psychologique générale aspécifique qui a pu être dénommé démoralisation (Frank, 1973) ou affect négatif (Tellegen, 1985 ; Clark et Watson, 1991).

L'affect négatif est envisagé en tant que facteur général de souffrance, non spécifique, et correspond au sentiment du sujet d'être triste, coupable, en colère, inquiet. L'absence d'affect négatif se retrouve dans des états de calme et de détente. L'affect positif — qui ne représente pas uniquement l'aspect inverse de l'affect négatif — correspond au sentiment d'être actif, intéressé, enthousiaste. Son absence est reflétée par des termes qui évoquent la fatigue et la langueur (Clark et Watson, 1991). L'anxiété correspondrait à un affect négatif élevé associé à un affect positif

normal alors que dans la dépression on retrouverait un affect négatif élevé mais un affect positif bas ou absent. Les relations de ces dimensions avec la personnalité s'avèrent également importantes (Clark et al., 1994).

LA COMORBIDITÉ DES TROUBLES ANXIEUX ET DÉPRESSIFS

Le recours aux critères opérationnalisés de diagnostic et l'utilisation d'instruments diagnostiques standardisés dans l'évaluation de la fréquence des troubles mentaux, non seulement au plan transversal mais également sur l'ensemble de la vie, a entraîné un intérêt accru dans les études de comorbidité. Celles-ci se définissent au sens épidémiologique comme les études de la fréquence et des interrelations des différents troubles mentaux qui peuvent survenir chez une même personne sur une période de temps définie (sur la vie entière, sur 6 mois, sur 1 mois) (Burke *et al.*, 1990; Maser et Cloninger, 1990).

Les études épidémiologiques récentes (Angst *et al.*, 1990; Boyd *et al.*, 1984; Degonda et Angst, 1993; Karno *et al.*, 1988; Kessler *et al.*, 1994; Lépine *et al.*, 1989; Schneider *et al.*, 1992; Wittchen et Von Zerssen, 1988; Wittchen *et al.*, 1992), ainsi que les études cliniques (Barlow *et al.*, 1986; Breier *et al.*, 1984; Di Nardo *et al.*, 1990; Leckman *et al.*, 1983; Lépine *et al.*, 1993a; Lesser *et al.*, 1988) ont fait état de taux élevés de comorbidité entre les troubles mentaux en général et les troubles dépressifs et anxieux en particulier. De plus, les études familiales ont retrouvé une forte association de la transmission de l'anxiété et de la dépression suggérant une vulnérabilité génétique commune et le rôle de facteurs environnementaux spécifiques dans l'expression des troubles (Kendler *et al.*, 1986; Merikangas *et al.*, 1994).

À titre d'exemple, les corrélations génétiques entre dépression majeure et anxiété généralisée seraient comprises entre 0,83 et 1,0 selon les niveaux de définition diagnostiques de l'anxiété généralisée suggérant des influences génétiques semblables pour la dépression et l'anxiété en terme de vulnérabilité générale alors que les facteurs environnementaux détermineraient les aspects cliniques spécifiques anxieux ou dépressifs (Kendler *et al.*, 1992).

Cependant, les taux de comorbidité varient considérablement d'une étude à l'autre, principalement dans les études transversales, ce qui suggère que des variables telles que la chronicité et la sévérité de la maladie, des différences au niveau du processus d'accès au soin et l'utilisation de procédures d'évaluation différente jouent un rôle certain.

Dans le cadre d'une étude internationale du *Composite International Diagnostic Interview* menée dans 18 centres à l'instigation de l'OMS et de l'*Alcohol, Drug and Mental Health Administration* (ADAMHA), nous avons évalué la comorbidité des troubles anxieux et affectifs chez des sujets vus dans des cadres différents (milieu psychiatrique spécialisé et médecine générale) (Lépine *et al.*, 1993b). Le principal résultat montrait que la présence d'un trouble anxieux et d'un trouble dépressif comorbides était de loin beaucoup plus fréquente que la présence d'un trouble unique ou d'une association de troubles anxieux entre eux ou de troubles dépressifs entre eux.

L'étude de cette comorbidité retrouve un ordre d'apparition des troubles dans lequel les troubles anxieux — et en particulier phobiques — précèdent la survenue des troubles dépressifs (Angst *et al.*, 1990; Lesser *et al.*, 1988; Schneier *et al.*, 1992;

Wittchen et Essau, 1993a). Dans le cadre du trouble panique et de l'anxiété généralisée, cependant une survenue la même année que l'épisode dépressif est le plus souvent observée dans environ 50 % des cas (Lépine *et al.*, 1993b).

Cette comorbidité est associée à une plus grande altération du fonctionnement social, un recours aux soins plus fréquent ainsi qu'une fréquence plus élevée des antécédents de conduites suicidaires retrouvées dans les études épidémiologiques (Weissman *et al.*, 1989 : Johnson *et al.*, 1990) et cliniques (Noyes *et al.*, 1991 ; Lépine *et al.*, 1993a).

LES TROUBLES ANXIEUX ET DÉPRESSIFS SUBSYNDROMIQUES

Comme le souligne Tyrer (1989), dans le domaine des troubles névrotiques, les symptômes mixtes anxieux et dépressifs sont de règle. Cet auteur a proposé une classification de l'anxiété et de la dépression selon la prédominance et la sévérité symptomatique (tableau 1).

Tableau 1. *Classification de l'anxiété et de la dépression. Prédominance des symptômes anxieux et/ou dépressifs (d'après Tyrer, 1989)*

Dépression		*Anxiété et dépression*		*Anxiété*	
Sévère	Léger	Sévère	Léger	Sévère	Léger
Mélancolie dépression « psychotique »	Dysthymie	Syndrome névrotique général	Névrose subclinique co-thymie	Agoraphobie (dont panique)	Trouble anxiété généralisée

Chez les patients présentant à la fois des symptômes anxieux et dépressifs, il différencie le syndrome névrotique général — forme la plus sévère — d'un état plus léger qu'il propose de nommer la névrose subclinique ou co-thymie.

Au plan épidémiologique, la prévalence des syndromes anxieux et dépressifs mixtes définis par la coexistence de troubles subsyndromiques anxieux et dépressifs semble peu élevée comparativement à celle des troubles bien définis — anxieux et dépressifs — ou des troubles subsyndromiques pris isolément. Wittchen et Essau (1993b) ont en effet retrouvé une prévalence de 0,8 %. Cependant, tout comme chez les sujets ayant une comorbidité de troubles anxieux et dépressifs bien définis, l'association de troubles subsyndromiques anxieux et dépressifs semble associée à un niveau de handicap et d'utilisation des soins plus élevé et un taux de rémission plus faible que lorsque les troubles surviennent isolément.

Dans une approche similaire, Judd *et al.* (1994) ont procédé à une réanalyse des données de l'enquête ECA et proposé de définir une nouvelle entité qu'ils ont dénommée *anxiété et dépression symptomatiques subsyndromiques* (*Subsyndromal Symptomatic Anxiety and Depression* – SSAD). Selon ces auteurs, cette entité correspondrait à la définition suivante : présence simultanée de deux symptômes ou plus de dépression et de deux symptômes anxieux ou plus chez un sujet qui ne remplit pas les critères de diagnostic d'un autre trouble affectif, anxieux ou somatoforme.

La prévalence de ce trouble sur un mois serait de 1 % lors de la première vague d'enquête et 1,5 % des sujets présenteraient un tel trouble lors d'une période de suivi

d'un an ce qui, au total, correspondrait à une prévalence sur un an de 2,5 % dans la population générale. Environ 23 % de ces sujets auraient recours à des soins médicaux ambulatoires, ils consulteraient en psychiatrie deux fois plus souvent que les sujets n'ayant aucun trouble mental au *Diagnostic Interview Schedule* et auraient ressenti au cours de leur vie des idées suicidaires, voire auraient fait plus fréquemment, mais de manière statistiquement non significative, des tentatives de suicide (3 % *vs* 1 %).

LE TROUBLE ANXIEUX ET DÉPRESSIF MIXTE

Compte tenu de l'association fréquente de symptômes anxieux et dépressifs, on peut supposer que de nombreux médecins qualifient les troubles des patients d'états anxio-dépressifs. La dernière révision de la CIM-10 a retenu une telle catégorie afin de tenir compte de l'existence de nombreux syndromes psychopathologiques ne satisfaisant pas aux critères actuels des classifications des troubles mentaux et supposés être vus principalement en médecine générale mais dans une perspective limitée et restrictive. En effet, selon la CIM-10, « cette catégorie doit être utilisée quand le sujet présente à la fois des symptômes anxieux et des symptômes dépressifs, sans prédominance nette des uns ou des autres, et sans que l'intensité des uns ou des autres ne soit suffisante pour justifier un diagnostic isolé. Quand des symptômes anxieux et dépressifs sont présents simultanément avec une intensité suffisante pour justifier des diagnostics isolés, les deux diagnostics doivent être notés et on ne fait pas un diagnostic de trouble anxieux et dépressif mixte. » On précise également la possibilité d'inclure dans ce cadre la dépression anxieuse lorsqu'elle demeure légère ou non persistante. Une telle description implique que cette nouvelle entité correspond davantage à un état subsyndromique avec des symptômes légers.

Au plan des critères diagnostiques pour la recherche, la CIM-10 donne les recommandations suivantes : « Les combinaisons possibles de symptômes relativement légers de ces troubles sont très nombreuses. Elles ne sont pas définies par des critères diagnostiques, autres que ceux donnés dans les *Descriptions cliniques et directives pour le diagnostic*. Les chercheurs souhaitant étudier des patients présentant l'un de ces troubles sont invités à établir leurs propres critères diagnostiques, à partir des directives pour le diagnostic, et en fonction des conditions et des objectifs de leurs études. »

Dans le contexte de la révision du DSM, au plan du DSM-IV, selon Barlow (1994), quatre options différentes étaient *a priori* envisagées : inclure ce trouble dans la classification dans la section des troubles de l'humeur ou des troubles anxieux, ne pas le mentionner, adapter les critères des troubles de l'humeur et des troubles anxieux (par exemple dysthymie ou dépression majeure et anxiété généralisée) pour tenir compte de ces sujets, ou l'inclure en annexe en tant que catégorie permettant des études ultérieures, position qui a été retenue comme nous le verrons.

L'ENQUÊTE DE TERRAIN DU DSM-IV

Une étude de terrain du trouble anxieux et dépressif mixte a été menée dans divers centres incluant cinq centres de santé primaire et deux consultations psychiatriques

ambulatoires (Zinbarg *et al.*, 1991; Zinbarg *et al.*, 1994). La batterie d'évaluation comprenait le *General Health Questionnaire* – GHQ (Goldberg et Hillier, 1979), le *Medical Outcome Study/Rand Short Form General Health Survey* – MOS (Stewart *et al.*, 1988; Wells *et al.*, 1989), l'*Anxiety Disorders Interview Schedule-Revised* – ADIS-R incluant le *Global Assessment of Current Functioning* (DiNardo et Barlow, 1988), les échelles d'Hamilton « anxiété et dépression » ainsi que leur version révisée (Riskind *et al.*, 1987) dans lesquelles, outre chaque item, chaque symptôme était coté séparément et le *Chronic Disease Score* ou CDS (Von Korff *et al.*, 1992) évaluant sur la base des traitements prescrits l'existence d'une éventuelle maladie chronique. Globalement, 666 patients ont été évalués. Les analyses ont porté principalement sur les sujets ayant un diagnostic de trouble anxieux et/ou dépressif non spécifié (NS) qui représentaient 13,7 % des patients vus en consultation psychiatrique et 8,0 % de ceux vus en centre de santé primaire. Les résultats ont montré que les cas présentant des symptômes affectifs et ne satisfaisant pas aux critères des troubles de l'axe I étaient aussi communs que les troubles anxieux ou dépressifs bien définis et étaient associés à un handicap. Parmi ces cas non spécifiés, 14,1 % avaient présenté un épisode dépressif majeur dans leurs antécédents.

À partir des symptômes cotés dans les versions révisées des échelles d'Hamilton, une analyse en composantes principales avec rotation varimax a été effectuée et une solution à trois composantes a été retenue. La première correspond à l'anxiété, la seconde à l'activation physiologique (troubles neurovégétatifs) et la troisième à la dépression; plusieurs symptômes saturant à la fois sur la composante anxiété et la composante dépression. Sur les bases de cette analyse, 4 échelles symptomatiques ont été construites :

– affect négatif de 14 items saturant sur les composantes anxiété et dépression;

– dépression de 6 items évaluant surtout l'anhédonie et une valeur basse de l'affect positif et saturant sur la composante dépression et peu sur l'anxiété;

– anxiété de 6 items comprenant la tension anxieuse et les préoccupations excessives et saturant sur la composante anxiété et peu sur la dépression;

– activation physiologique de 9 items saturant relativement peu sur les composantes anxiété et dépression.

Le groupe des cas NS était alors caractérisé par un score élevé à l'échelle d'affect négatif et l'absence d'élévation nette sur les trois autres échelles, des scores plus élevés que ceux cotés chez les patients n'ayant pas de trouble mental, un score à l'échelle d'anxiété plus bas que celui des patients ayant une anxiété généralisée, un score à l'échelle de dépression plus bas que celui des patients ayant un épisode dépressif majeur et un score à l'échelle d'activation physiologique également plus bas que celui des patients ayant un trouble panique ou une agoraphobie. Finalement, à partir des 14 items de l'échelle d'affect négatif, en regroupant les troubles du sommeil en un seul item ainsi que les troubles cognitifs et en éliminant l'item culpabilité, dont la corrélation avec le score total était faible, on obtient une échelle de 10 items. En utilisant un seuil de 4 symptômes ou plus à cette échelle et en excluant les sujets satisfaisant actuellement ou dans le passé aux critères d'épisode dépressif majeur, d'anxiété généralisée ou de dysthymie ainsi qu'à ceux de tout autre trouble actuel de l'axe I, sauf la phobie simple, 54 % des cas NS recevraient un diagnostic de trouble anxieux et dépressif mixte.

Tableau 2. *Critères de recherche pour le trouble mixte anxiété-dépression (DSM-IV)*

A. Humeur dysphorique persistante (ou récurrente) depuis au moins un mois.

B. L'humeur dysphorique est, depuis au moins un mois, associée à quatre symptômes (au moins) parmi les suivants :

(1) difficultés de concentration ou impression de « tête vide »,
(2) troubles du sommeil (difficultés à s'endormir ou à rester endormi ou sommeil non reposant, non satisfaisant),
(3) fatigue ou baisse d'énergie,
(4) irritabilité,
(5) préoccupations, soucis
(6) pleurs faciles,
(7) hypervigilance,
(8) anticipation du pire,
(9) sans espoir (pessimisme envahissant concernant le futur),
(10) baisse de l'estime de soi ou dévalorisation.

C. Les symptômes induisent une souffrance clinique significative ou une altération du fonctionnement social, professionnel ou dans d'autres domaines importants.

D. Les symptômes ne sont pas imputables aux effets physiologiques directs d'une substance (ex. une substance donnant lieu à abus, un médicament) ou d'une affection médicale générale.

E. Tous les critères suivants sont réunis :

(1) N'a jamais rempli les critères d'un trouble dépressif majeur, d'un trouble panique, d'un trouble anxiété généralisé.
(2) Ne répond pas actuellement aux critères des autres troubles anxieux ou dépressifs (y compris trouble anxieux ou dépressif en rémission partielle).
(3) Les symptômes ne sont pas imputables à un autre trouble mental.

Des critères de recherche ont donc été proposés en annexe du DSM-IV (tableau 2). Ils requièrent la présence d'une humeur dysphorique persistante (ou récurrente), accompagnée de la présence d'au moins 4 symptômes de cette liste de 10 symptômes (correspondant à la dimension de l'affect négatif), entraînant une souffrance marquée ou un handicap significatif et, surtout, survenant chez un patient dont les éléments cliniques ne remplissent pas actuellement les critères d'aucun autre trouble de l'axe I (sauf la phobie simple), et n'ayant jamais satisfait aux critères d'une dépression majeure, d'une dysthymie ou d'un trouble anxiété généralisée.

CONCLUSION

L'existence d'une association fréquente entre anxiété et dépression est largement reconnue. Cependant, cette relation a conduit à la formulation de modèles théoriques et conceptuels très variés. Au cours de ces dernières années, l'accent a été mis sur l'importance des troubles subsyndromiques qui seraient fréquemment observés en médecine générale. Certains demeurent partisans d'une approche purement dimen-

sionnelle dans l'étude de ces manifestations psychopathologiques. Pour d'autres, il convient d'intégrer, dans une perspective catégorielle, ces phénomènes dans les classifications actuelles. Le danger que constitue le recours à des catégories de dénomination incertaine et peu définies ne doit pas cependant être minimisé. La catégorie du trouble anxieux et dépressif mixte est récente et n'a pas fait l'objet de travaux adéquats au plan de sa validité, de son évolution et de son traitement. Elle a été intégrée dans la CIM-10 mais ne figure pas dans la classification du DSM-IV ou elle ne constitue qu'une proposition pouvant faire l'objet de travaux de recherche. En particulier, on ne sait pas dans quelle mesure ces états peuvent correspondre à des phases prodromiques ou résiduelles d'autres troubles plus sévères et bien définis.

Indéniablement, de nombreuses stratégies de recherche s'avèrent nécessaires. Selon Katon et Roy-Byrne (1991), il convient de définir les critères opérationnels, le nombre, le type et la durée des symptômes en précisant clairement le niveau du handicap fonctionnel; à ce titre, l'étude de terrain a apporté des premiers éléments de réponse. Il convient également d'entreprendre des études longitudinales pour savoir quel est le risque de survenue de troubles anxieux et dépressifs plus sévères voire s'il existe un sous-groupe non chronique qui s'améliore sans traitement (que Katon et Roy-Byrne proposent d'appeler trouble de l'adaptation avec manifestations émotionnelles mixtes), de connaître le rôle des facteurs de stress, des troubles de la personnalité, des mécanismes d'adaptation, de l'estime de soi et du support social chez ces patients et enfin de préciser les indications — si nécessaire — des stratégies thérapeutiques psychopharmacologiques et psychothérapiques ainsi que leur durée.

BIBLIOGRAPHIE

AMERICAN PSYCHIATRIC ASSOCIATION — Diagnostic and Statistical Manual of Mental Disorders, 4th edition = DSM-IV, Washington, 1994. Traduction française sous la direction de J. D. GUELFI *et al.*, Masson, Paris, sous presse.

ANGST J., VOLLRATH M., MERIKANGAS K. R., ERNST C. — Co-morbidity of Anxiety and Depression in the Zurich Cohort Study of Young Adults. In : MASER J. D. et CLONINGER C. R. (Eds), *Co-morbidity of Mood and Anxiety Disorders,* 123-153, Washington D.C., American Psychiatric Press, 1990.

BARLOW D. H. — *Mixed Anxiety/Depression : DSM-IV Field Trial Data.* 147th Annual Meeting of the American Psychiatric Association, Philadelphie, 1994.

BARLOW D. H., DINARDO P. A., VERMILYEA B. B., VERMILYEA J. A., BLANCHARD E. — Co-morbidity and Depression among the Anxiety Disorders : Issues in Classification and Diagnosis. *Journal of Nervous and Mental Disease*, *174*, 63-72, 1986.

BARRETT J. E., BARRETT J. A., OXMAN T. E., GERBER P. D. — The Prevalence of Psychiatric Disorders in General Practice. *Archives of General Psychiatry*, *45*, 1100-1106, 1988.

BOYD J. H., BURKE J. D., GRUENBERG E., HOLZER C. E. III, RAE D. S., GEORGE L. K., KARNO M., STOLTZMAN R., MCEVOY L., NESTADT G. — Exclusion Criteria of DSM-III : a Study of Co-occurrence of Hierarchy-free Syndromes. *Archives of General Psychiatry*, *41*, 983-989, 1984.

BREIER A., CHARNEY D. S., HONINGER G. R. — Major Depression in Patients with Agoraphobia and Panic Disorder. *Archives of General Psychiatry*, *41*, 1129-1135, 1984.

BURKE J. D., WITTCHEN H. U., REGIER D. A., SARTORIUS N. — Extracting Information from Diagnostic Interviews on Co-occurrence of Symptoms of Anxiety and Depression. In :

J. D. MASER et C. R. CLONINGER (Eds), *Co-morbidity of Mood and Anxiety Disorders*, American Psychiatric Press, Washington D.C., 1990.

CLARK L. A. — The Anxiety and Depressive Disorders : Descriptive Psychopathology and Differential Diagnosis. In : P. C. KENDALL et D. WATSON (Eds), *Anxiety and Depression : Distinctive and Overlapping Features*, Academic Press, New York, 83-129, 1989.

CLARK L. A., WATSON D. — Tripartite Model of Anxiety and Depression : Psychometric Evidence and Taxonomic Implications, *Journal of Abnormal Psychology*, *100*, 316-336, 1991.

CLARK L. A., WATSON D., MINEKA S. — Temperament, Personality, and the Mood and Anxiety Disorders, *Journal of Abnormal Psychology*, *103*, 103-116, 1994.

DEGONDA M., ANGST J. — The Zurich Study. XX. Social Phobia and Agoraphobia. *European Archives of Psychiatry and Clinical Neurosciences*, *243*, 95-102, 1993.

DINARDO P. A., BARLOW D. H. — *Anxiety Disorders Interview Schedule-Revised (ADIS-R)*, Albany : Center for Stress and Anxiety Disorders, University at Albany, State University of New York, 1988.

DINARDO P. A., BARLOW D. H. — Syndrom and Symptom Co-occurrence in the Anxiety Disorders. In : MASER J. D. et CLONINGER C. R. (Eds). *Co-morbidity of Mood and Anxiety Disorders*, American Psychiatric Press, Washington D.C., 205-230, 1990.

FEINSTEIN A. R. — The Pre-Therapeutic Classification of Co-morbidity in Chronic Disease. *Journal of Chronic Diseases*, *23*, 455-468, 1970.

FRANK J. D. — *Persuasion and Healing*. John Hopkins University Press, Baltimore, 1973.

GOLDBERG D., HILLIER V. C. — A Scaled Version of the General Health Questionnaire. *Psychological Medicine*, *9*, 139-145, 1979.

JOHNSON J., WEISSMAN M. M., KLERMAN G. L. — Panic Disorder, Co-morbidity, and Suicide Attempts. *Archives of General Psychiatry*, *47*, 805-808, 1990.

JUDD L. L., RAPAPORT M. H., PAULUS M. P., BROWN J. L. — Subsyndromal Symptomatic Depression : a New Mood Disorder? *Journal of Clinical Psychiatry*, *55*, Suppl : 4, 1994.

KARNO M., GOLDING J. M., SORENSON S. B., BURNAM M. A. — The Epidemiology of Obsessive-Compulsive Disorder in Five US Communities. *Archives of General Psychiatry*, *45*, 1094-1099, 1988.

KATON W., ROY-BYRNE P. — Mixed Anxiety and Depression. *Journal of Abnormal Psychology*, *100*, 3, 337-345, 1991.

KENDLER K. S., HEATH A., MARTIN N. G. — Symptoms of Anxiety and Depression in a Volunteer Twin Population. *Archives of General Psychiatry*, *43*, 213-221, 1986.

KENDLER K. S., NEALE M. C., KESSLER R. C., HEALTH A. C., EAVES L. J. — Major Depression and Generalized Anxiety Disorder : Same genes, (partly) Different Environments? *Archives of General Psychiatry*, *49*, 716-722, 1992.

KESSLER R. C., MCGONAGIE K. A., ZHAO S., NELSON C. B., HUGHES M., ESHLEMAN S., WITTCHEN H. U., KENDLER K. S. — Lifetime and 12-Month Prevalence of DSM-III-R Psychiatric Disorders in the United States : Results from the National Co-morbidity Survey. *Archives of General Psychiatry*, *51*, 8-19, 1994.

LECKMAN J. F., WEISSMAN M. M., MERIKANGAS K. R., PAULS D. L., PRUSOFF B. A. — Panic Disorder and Major Depression. *Archives of General Psychiatry*, *40*, 1055-1060, 1983.

LÉPINE J. P., CHIGNON J. M., TÉHÉRANI M. — Suicide Attempts in Patients with Panic Disorder. *Archives of General Psychiatry*, *50*, 144-149, 1993a.

LÉPINE J. P., LELLOUCH J., LOVELL A., TÉHÉRANI M., HA C., VERDIER-TAILLEFER M. H., RAMBOURG N., LEMPÉRIÈRE T. — Anxiety and Depressive Disorders in a French Population : Methodology and Preliminary Results. *Psychiatry and Psychobiology*, *4*, 267-274, 1989.

LÉPINE J. P., WITTCHEN H. U., ESSAU C. A. and participants in the multicentre WHO/ADAMHA field trial. Lifetime and Current Co-morbidity of Anxiety and Affective Disorders : Results from the International WHO/ADAMHA CIDI Field Trials. *International Journal on Methods in Psychiatric Research*, 67-77, 1993b.

LESSER I. M., RUBIN R. T., PECKNOLD J. C., RIFKIN A., SWINSON R. P., LYDIARD R. B., BURROWS G. D., NOYES, Jr R., DUPONT R. L. — Secondary Depression in Panic Disorder and Agoraphobia : I. Frequency, Severity and Response to Treatment. *Archives of General Psychiatry*, *45*, 437-443, 1988.

MASER J. D., CLONINGER C. R. (Eds) — *Co-morbidity of Mood and Anxiety Disorders.* American Psychiatric Press, Washington D.C.

MERIKANGAS K. R., RISCH N. J., WEISSMAN M. M. — Co-morbidity and Cotransmission of Alcoholism, Anxiety and Depression. *Psychological Medicine*, *24*, 69-80, 1994.

NOYES R. — Suicide and Panic Disorder : a Review. *Journal of Affective Disorders*, *22*, 1-11, 1991.

RISKIND J H., BECK A. T., BROWN G., STEER R. A. — Taking the Measure of Anxiety and Depression : Validity of the Reconstructed Hamilton Rating Scales. *Journal of Nervous and Mental Disease*, *175*, 474-479, 1987.

SCHNEIER F. R., JOHNSON J., HORNIG C. D., LIEBOWITZ M. R., WEISSMAN M. M. — Social Phobia. Co-morbidity and Morbidity in an Epidemiologic Sample. *Archives of General Psychiatry*, *49*, 282-288, 1992.

STEWART A., HAYS R., WARE J. — The MOS Short-Form of the General Health Survey. *Medical Care*, *26*, 724-735, 1988.

TELLEGEN A. — Structures of Mood and Personality and their Relevance to Assessing Anxiety, with an Emphasis on Self-Report in A. H. TUMA et J. D. MASER (Eds). *Anxiety and the Anxiety Disorders.* Hillsdale, NJ, Erlbaum, 681-706, 1985.

TYRER P. — *Classification of Neurosis.* John Wiley & Sons, Chichester, 1989.

VON KORFF M., WAGNER E. H., SAUMDERS K. — A Chronic Disease Score for Automated Pharmacy Data. *Journal of Clinical Epidemiology*, *45*, 197-203, 1992.

WEISSMAN M., KLERMAN G., MARKOWITZ J., OUELLETTE R. — Suicide Ideation and Suicide Attempts in Panic Disorder and Attacks. *New England Journal of Medicine*, *321*, 1209-1213, 1989.

WELLS K., STEWART A., HAYS R., BURNAM A., ROGERS W., DANIELS M., BERRY S., GREENFIELD S., WARE J. — The Functioning and Well-Being of Depressed Patients. *Journal of the American Medical Association*, *262*, 914-919, 1989.

WITTCHEN H. U., VON ZERSSEN D. (Eds) — *Verläufe behandelter und unbehandelter Depressionen und Angststörungen – Eine klinisch psychiatrische und epidemiologische Verlaufsuntersuchung.* Springer, Heidelberg, 1988.

WITTCHEN H. U., ESSAU C. A. — Epidemiology of Anxiety Disorders. In : *Psychiatry* (MICHELS R. Ed.), Lippincott J. B., Philadelphie, 1, 31.1, 1-25, 1993a.

WITTCHEN H. U., ESSAU C. A. — Co-morbidity and Mixed Anxiety-Depressive Disorders : Is there Epidemiological Evidence? *The Journal of Clinical Psychiatry*, 54, Suppl., 9-15, 1993b.

WITTCHEN H. U., ESSAU C. A., VON ZERSSEN D., KRIEG J. C., ZAUDIG M. — Lifetime and Six-month Prevalence of Mental Disorders in the Munich Follow-Up Study. *European Archives of Psychiatry and Clinical Neurosciences*, *241*, 247-258, 1992.

ZINBARG R. E., BARLOW D. H. — Mixed Anxiety-Depression : a New Diagnostic Category? In : R. M. RAPEE et D. H. BARLOW (Eds), *Chronic Anxiety – Generalized Anxiety Disorder and Mixed Anxiety-Depression.* Guilford Press, New York, 136-152, 1991.

ZINBARG R. E., BARLOW D. H., LIEBOWITZ M., STREET L., BROADHEAD E., KATON W., ROY-BYRNE P., LÉPINE J. P., TÉHÉRANI M., RICHARDS J., BRANTLEY J. P., KRAEMER H. — The DSM-IV Field Trial for Mixed Anxiety Depression. *Am. J. Psych. 151*, 1153-1162, 1994.

12

THÉRAPIES COGNITIVES : ÉVALUATION QUANTITATIVE DES DYSFONCTIONNEMENTS

par C. MIRABEL-SARRON

Les thérapies cognitives apparaissent vers les années 1960 et se proposent d'apporter une réponse thérapeutique aux troubles émotionnels anxieux ou dépressifs (Beck, 1976). Le modèle théorique empirique postule que le trouble de l'humeur est la conséquence de dysfonctionnements cognitifs. Ce point de vue est radicalement opposé aux travaux antérieurs qui concluaient que le trouble de l'affect était déterminant.

Dans la théorie cognitive, le dysfonctionnement cognitif existe sur trois niveaux qui sont appelés les cognitions, les processus cognitifs et les schémas cognitifs (Beck, 1976). Le niveau cognitif perturbé en premier lieu est celui du schéma. Le schéma est une structure psychologique développée dans l'enfance au cours d'expériences. L'individu possède plusieurs schémas qui, avec la maturité, vont se multiplier et se complexifier. Certains de ces schémas sont dits dépressogènes parce qu'ils représentent une vulnérabilité à la dépression. Envisageons leurs caractéristiques : ils se présentent comme des règles morales, rigides, inflexibles et ils concernent des aspects de la personne en rapport avec les jugements qu'elle a sur elle-même et dans ses relations interpersonnelles. Les schémas sont idiosyncrasiques et sont activés par des stimuli internes ou externes analogues aux circonstances qui sont à l'origine de la construction du schéma. Au travers de son schéma, l'individu code, organise et évalue les différents événements qu'il perçoit. Il s'agit par conséquent d'une lecture de la réalité au travers d'un « filtre personnel ». Cette première organisation des informations va interagir avec d'autres activités cognitives.

Les processus cognitifs sont des mécanismes de pensée qui interviennent dans le raisonnement. Il est classique de parler de processus inductifs, déductifs et analogiques. Beck propose à partir de son expérience clinique une classification originale de ces processus. Il distingue les processus principaux (surgénéralisation, abstraction sélective, personnalisation), les processus secondaires (maximalisation, minimisation), les erreurs de style (dichotomie) et les erreurs sémantiques. Ces processus existeraient chez tout individu. Cependant, chaque entité psychopathologique se caractériserait par une modification dans la fréquence d'emploi de ces processus (Beck, 1976). L'information issue du traitement cognitif par les schémas et les processus se présente sous la forme de cognitions. Les cognitions sont les pensées non volontaires, brèves, précises, plausibles qui traversent brièvement l'esprit du

sujet et précèdent le mouvement émotionnel. Elles ont un contenu négatif chez le sujet déprimé et leur forme syntaxique reflète la rigidité du schéma («*je dois donner chaque jour un sens à ma vie*», «*je dois plaire à tout le monde*», «*il faut toujours être à l'écoute des autres*», «*je suis un raté*»,...).

Calquées sur ce modèle théorique, les thérapies cognitives abordent cognitions, processus et schémas de façon à ce que l'individu apprenne à repérer ses raisonnements dysfonctionnels, puis à substituer d'autres interprétations de la réalité puis à identifier la vulnérabilité due au schéma activé.

Afin d'évaluer ces différents niveaux de dysfonctionnements cognitifs, des instruments ont été construits. Ce sont principalement des questionnaires qui ont pour objectif d'évaluer les altérations des cognitions, des processus cognitifs et des schémas. Ces outils d'évaluation se sont développés dans le domaine de la dépression et de l'anxiété. Le tableau 1 présente l'ensemble de ces questionnaires.

Du fait que seuls les questionnaires intéressant la dépression ont été traduits puis validés dans leur version française, nous choisirons d'abord spécifiquement ce domaine.

Dans une première partie, nous aborderons les questionnaires les plus utilisés, avec leurs spécificités et leurs limites. Dans une deuxième partie, nous envisagerons les différentes utilisations de ces questionnaires.

LES INSTRUMENTS ET LEURS CARACTÉRISTIQUES

Notre présentation favorisera l'étude de quatre questionnaires qui sont les plus utilisés, aussi bien dans une pratique clinique que dans les études internationales. Ce sont l'inventaire pour la dépression de Beck (BDI), l'échelle des pensées automatiques (ATQ), l'échelle de désespoir (HS) et l'échelle des attitudes dysfonctionnelles (DAS). Le tableau 1 présente, par niveau de dysfonctionnement cognitif, les différents questionnaires.

Tableau 1. *Évaluations cognitives dans la dépression*

Niveau des cognitions	
– Inventaire pour la dépression de Beck (BDI)	1961
– *Irrational Belief Test* (IBT) de Jones	1969
– Échelle de désespoir (HS) de Beck	1974
– Échelle des pensées automatiques (ATQ) de Hollon et Kendall	1980
– *Crandell Chambless Inventory* (CCI)	1981
– *Cognitive Check List* (CCL) de Beck	1988
Niveau des processus cognitifs	
– *Cognitive Error Questionnaire* (CEQ) de Lefebvre	1981
Niveau des schémas cognitifs	
– Échelle des attitudes dysfonctionnelles (DAS) de Weissman	1978
– Échelle sociotropie autonomie (SAS) de Beck	1986
– *Beck Self-Concept test*	1990

Niveau des cognitions

Dès la formulation du modèle théorique, il est apparu nécessaire de quantifier le dysfonctionnement cognitif au niveau où il était le plus abordable, celui des cognitions encore appelées pensées automatiques. Les chercheurs ont essayé de distinguer les cognitions spécifiques à l'humeur dépressive, puis ont voulu estimer leur fréquence d'apparition et la signification clinique qui pouvait y être apportée.

L'inventaire pour la dépression de Beck (Beck, Ward, Mendelson, Mock, Erbaugh, 1961)

Il s'agit d'un auto-questionnaire qui permet d'évaluer l'intensité de la symptomatologie subjective de la dépression.

Pour la construction de l'échelle, les auteurs ont recueilli un ensemble d'attitudes et de symptômes qui semblent spécifiques des déprimés et qui correspondent à la description de la dépression dans la littérature.

Le questionnaire initial comprend 21 catégories de symptômes. Pour chaque catégorie, il est proposé une série graduée de 4 propositions présentées selon le degré de sévérité (de 0 à 3 maximum de sévérité).

Le score total au questionnaire s'étend de 0 à 63 et reflète la profondeur subjective de la dépression.

Deux études de validation montrent une forte corrélation entre la dépression estimée cliniquement sur 4 indices et le score au BDI. Ainsi, dans sa version anglaise, le BDI permet de discriminer des groupes de patients dont l'intensité de la dépression est différente. Il permet de plus de refléter des modifications de l'intensité de la dépression chez un même malade.

La traduction et la validation françaises reviennent à Pichot, Lempérière et Mirouze (1963).

La passation du BDI s'effectue chez 69 sujets déprimés. Il est alors constaté que deux catégories symptomatiques ont des corrélations plus faibles (F/ perte de poids; et K/ irritabilité). Cette constatation avait été faite par Beck pour le facteur F, dans la version anglaise. C'est-à-dire que la perte de poids n'est pas proportionnelle à la gravité du syndrome dépressif et que l'irritabilité n'est pas une donnée homogène. En revanche, les 4 items qui ont la corrélation la plus forte avec le facteur général sont : le pessimisme, l'indécision, l'insatisfaction et la haine de soi-même.

La cotation du questionnaire fournit, pour valeurs seuils, dans la version anglaise, les données suivantes :

- de 0 à 10,9 : pas de dépression,
- de 11 à 18,7 : dépression légère,
- de 18,8 à 25,4 : dépression moyenne,
- au-delà de 25,5 : dépression sévère.

Dans la version française, Pichot obtient, pour valeur moyenne du BDI chez les 69 sujets déprimés, une valeur de 27,01.

Dans l'étude française, l'échantillon regroupait des dépressions endogènes, non endogènes et d'involution. Le score au BDI ne permet pas de distinguer ces différents groupes nosologiques.

Plus tardivement apparaissent deux autres versions du BDI, une version abrégée en 13 items (Beck, Beamesberfer, 1974) et une version étendue en 25 items « Beck-Pichot » (Bobon, Sanchez-Blaque, von Frenckell, 1981).

Une traduction française de l'échelle abrégée est proposée par Collet et Cottraux en 1986. Cette version est plutôt destinée aux généralistes et aux chercheurs pour effectuer une évaluation rapide de la dépression. Cette échelle abrégée n'a retenu que les items les plus fortement corrélés avec la note globale du BDI 21, et ceux dont la corrélation était maximale avec l'évaluation des cliniciens. Le questionnaire en 13 items corrèle à 0,96 avec un score total de l'échelle originale et 0,61 avec l'évaluation des cliniciens.

En France, plusieurs études ont été effectuées. En 1984, Lempérière *et al.* l'utilisent chez 79 déprimés sévères qui obtiennent un score moyen de 22 au BDI 13. Cette échelle est corrélée significativement à l'échelle de Hamilton pour la dépression et plus encore à l'échelle de Montgomery et Asberg (MADRS).

Collet et Cottraux en 1986 proposent le BDI 13 à 50 patients déprimés vus en consultation; la note moyenne obtenue est 18 plus ou moins 6.

Les auteurs mettent en évidence une corrélation avec l'échelle de Hamilton pour la dépression (HDRS) et l'échelle de ralentissement de Widlocher (ERD).

L'échelle comprend 13 catégories symptomatiques (de A à M). Son score total varie de 0 à 39. Les seuils de sévérité retenus sont ceux définis par Beck et Beamesberfer (1974) :

- de 0 à 4 : pas de dépression,
- de 5 à 7 : dépression légère,
- de 8 à 15 : dépression modérée,
- 16 et plus : dépression sévère.

La forme étendue en 25 items est une construction élaborée à partir de la forme abrégée du BDI à laquelle Pichot adjoint 12 items pour explorer d'autres syndromes de dépression comme celui de l'anxiété.

Au total, l'inventaire pour la dépression de Beck est largement utilisé depuis plus de 25 années (Beck, Steer, Garbin, 1988). Sa version en 21 items est la forme classique utilisée dans les études de recherche. De passation simple, il permet une évaluation des aspects subjectifs de la dépression.

L'échelle de désespoir ou HS (Beck, Weissman, Lester, Treixler, 1974)

Il s'agit du deuxième questionnaire construit par Beck et ses collaborateurs en 1974; il a pour objectif de mesurer l'intensité du pessimisme dans différents contextes psychopathologiques (dépression, alcoolisme, schizophrénie, psychopathie,...). Le pessimisme serait un système de prédictions négatives concernant soi-même et le futur. La construction de la version anglaise du questionnaire fait appel à deux sources : 9 items ont été sélectionnés à partir d'un test d'attitude sur le futur et 11 items ont été sélectionnés par des cliniciens dans le discours de sujets déprimés désespérés. Cette échelle se compose au total de 20 items. Le score total représente l'intensité du pessimisme et est prédictif du risque suicidaire.

La traduction et la validation de l'échelle reviennent à Bouvard, Charles, Guerin, Aimard, Cottraux (1992).

L'analyse factorielle met en évidence trois facteurs :

– sentiment négatif vis-à-vis du futur : 41,5 % de la variance exprimée (v.e.) ;
– perte de motivation : 6,2 % de la v.e. ;
– attente vis-à-vis de l'avenir : 5,6 % de la v.e.

Le score total de l'échelle est corrélé à l'intentionnalité suicidaire plutôt qu'à l'intensité de la dépression. Un score moyen de 13 (total maximum 20) a été retrouvé chez des sujets venant de faire une tentative de suicide.

La validité concourante de cette échelle est peu satisfaisante avec le BDI en 13 items, et non satisfaisante avec l'échelle de Hamilton pour la dépression (HDRS) ; elle est cependant satisfaisante avec l'échelle des pensées automatiques (ATQ).

En somme, ce questionnaire de passation simple permet une évaluation du risque suicidaire ; plus le score total est élevé et plus le risque suicidaire est grand.

L'échelle des pensées automatiques (ATQ)

Il s'agit d'un auto-questionnaire de 30 items destiné à mesurer la fréquence d'occurrence de pensées automatiques négatives associées à la dépression. La version anglaise du questionnaire (Hollon et Kendall, 1980) succède aux travaux de Jones (*Irrational Beliefs Test*, 1969). L'ATQ est construit à partir des récits d'épisodes dysphoriques fournis par 100 étudiants.

Les 30 items sélectionnés permettent de distinguer les populations dépressives des non dépressives. Aucune différence du contenu de pensée n'est retrouvée entre les hommes et les femmes. La traduction et la validation française reviennent à Bouvard, Charles, Guerin, Aimard et Cottraux (1992). Le sujet évalue pour chacune des 30 pensées automatiques proposées, la fréquence d'occurrence dans la semaine précédente, sur une échelle en six points.

Le score de l'échelle s'étend de 30 à 180. Le score moyen dans une population de déprimés est de 79,64 plus ou moins 22,29. Dans une population contrôle, une note moyenne de 48,57 plus ou moins 10,89 est retrouvée.

L'analyse factorielle met en évidence quatre facteurs :

– le désir de changement : 45,9 % de la v.e. ;
– les attentes négatives : 5,2 % de la v.e. ;
– la perte de l'estime de soi : 4,4 % de la v.e. ;
– le désespoir : 3,5 % de la v.e.

Ce questionnaire permet une évaluation quantitative du dysfonctionnement cognitif en mesurant le taux de fréquence des cognitions témoignant de l'hyperactivité cognitive (Beck, 1978). De plus, ce questionnaire permet d'estimer le changement des contenus des cognitions avec le temps. Cette mesure permettrait ainsi d'estimer les changements thérapeutiques. Une validité concourante satisfaisante est retrouvée avec le BDI.

Eaves et Rush (1984) montrent dans une étude contrôlée que le score global ATQ n'est pas significativement différent entre sujets témoins et sujets déprimés en rémission clinique. Les auteurs montrent également que le score ATQ est similaire chez les déprimés endogènes et non endogènes en phase dépressive. De ce fait, cette mesure cognitive est liée à l'état dépressif lui-même et l'amélioration clinique s'accompagne

d'une diminution de la fréquence d'apparition des cognitions négatives. Ces constatations sont en accord avec les théories de Beck (1976); la persistance d'une fréquence élevée de cognitions négatives serait prédictive d'une dépression traînante.

Cognitions Check List

Il s'agit du dernier inventaire des cognitions élaboré par Beck, Brown, Eidelson, Steer, Riskind (1987). Il se distingue des outils précédents par son intention d'évaluer la fréquence des cognitions négatives et positives non seulement dans la dépression mais aussi dans les troubles anxieux.

La CCL comprend 26 items : 14 sont consacrés à des cognitions pertinentes à l'état dépressif et 12 items concernent des cognitions cohérentes avec les troubles anxieux.

Effectivement, les thèmes cognitifs dépressifs sont l'auto-dépréciation et l'attitude négative envers le passé et le futur. Les cognitions anxieuses sont caractérisées par les thèmes de danger et la surestimation d'un accident dans le futur.

Cette échelle permet une discrimination correcte de l'anxiété et de la dépression d'après les cognitions. Ce fait est en accord avec les théories de Beck (1976).

Niveau des processus cognitifs

Les processus cognitifs trop fréquemment utilisés par les sujets déprimés sont tout d'abord la personnalisation puis la surgénéralisation et l'abstraction sélective. Habituellement, le clinicien repère par expérience l'emploi trop général de ces modes de raisonnement.

Un seul questionnaire élaboré par Lefebvre en 1981 évalue la quantité d'erreurs logiques dans le raisonnement du patient.

Le *Cognitive Error Questionnaire* (CEQ) est composé de 24 historiettes qui représentent des situations courantes. Il est demandé au sujet d'estimer comment il réagirait dans la même situation.

Ce questionnaire évalue quatre types de processus : la personnalisation, la surgénéralisation, l'abstraction sélective et la dramatisation. Le score global indique le niveau de distorsion logique du raisonnement. Ce questionnaire est rarement utilisé dans les travaux de recherche. Il n'existe pas de version française validée.

Niveau des schémas cognitifs

L'évaluation des schémas cognitifs revêt une importance théorique fondamentale étant donné leur rôle supposé dans la survenue de l'épisode dépressif. L'hypothèse est que le schéma est un trait de vulnérabilité idiosyncrasique. Un questionnaire explore principalement ce domaine : c'est l'échelle des attitudes dysfonctionnelles (DAS).

Échelle des attitudes dysfonctionnelles

Cet auto-questionnaire de 40 items est construit par Weissman en 1978 pour faciliter et systématiser l'identification des schémas cognitifs. Les croyances proposées

sont des idées de perfectionnisme, de réussite vis-à-vis du monde, de l'influence du jugement des autres,... Le score total du questionnaire s'échelonne entre 40 et 280.

La traduction et la validation française en reviennent à Bouvard *et al.* (1994).

Le score total de la DAS diminue avec l'amélioration symptomatologique de la dépression. Beck retient que dans une étude contrôlée (Kovaks, Rush, Beck, Hollon, 1981), la persistance d'un taux bas à la DAS chez les patients ayant bénéficié d'une thérapie cognitive fait la preuve de l'action de la thérapie sur des schémas cognitifs.

De telles conclusions sont remises en question par de nombreux auteurs qui estiment que la DAS n'évalue pas un trait mais plutôt un état. Eaves et Rush (1984) montrent que le score DAS s'améliore avec la rémission de la dépression. Cependant, les déprimés en rémission présentent un score DAS légèrement plus élevé que les sujets contrôle.

Silverman, Eardley (1984) obtiennent des résultats similaires et concluent que ces croyances sont liées à l'état. Silverman fait passer la DAS à d'autres groupes psychopathologiques (schizophrènes,...) qui présentent des scores DAS très élevés, supérieurs à ceux du groupe des déprimés majeurs. Il en conclut que les déprimés en rémission présentent un mode de pensée moins mal adapté.

Échelle autonomie-sociotropie

Ce questionnaire élaboré par Beck *et al.* en 1986 a pour objectif de repérer des traits d'autonomie ou de sociotropie supposés être des vulnérabilités à la dépression. La sociotropie est le besoin pour l'individu d'avoir des encouragements, de l'attention de la part des autres. L'autonomie est la nécessité pour un sujet de se centrer sur la réalisation de ses objectifs sans contrainte de la part des autres. Une personne autonome par exemple peut devenir dépressive si des facteurs internes ou externes sont perçus comme d'irréversibles obstacles à ses objectifs.

La version anglaise du questionnaire est élaborée à partir de propos de sujets déprimés en cours de thérapie cognitive. Le questionnaire final composé de 30 items autonomes et 30 items sociotropes présente une structure factorielle comprenant trois facteurs dans chacune des deux dimensions (A1 individualisme, A2 se libérer du contrôle, A3 préférence pour la solitude, S1 inquiétude par rapport à la désapprobation, S2 inquiétude par rapport à la séparation, S3 souci de plaire aux autres).

Plusieurs études confirment une corrélation entre S1, S2, S3 et le BDI 21. Une pareille corrélation n'est pas retrouvée avec la dimension autonomie.

Le Beck-Self-Concept Test proposé en 1990 est une nouvelle procédure d'évaluation (Segal, 1994)

Les autres instruments d'évaluation

Nous avons privilégié l'étude des questionnaires les plus souvent cités dans les publications internationales. Cependant, d'autres échelles et questionnaires ont été élaborés.

D'autres théories cognitives de la dépression existent comme celle de Seligman, dite de l'impuissance apprise, révisée en 1988 par Abramson. Cette théorie cognitive postule que la dépression fait suite à l'exposition à des stimulations douloureuses incontrôlables. Cette théorie élaborée, puis démontrée expérimentalement chez

l'animal, fut généralisée à l'être humain. Elle intègre alors des concepts de la psychologie sociale concernant les styles attributionnels. Le déprimé présente des attributions causales stables, globales et internes des événements négatifs. En revanche, ses attributions causales des événements heureux sont instables, spécifiques et externes.

Seligman propose en 1979 l'*Attribution Style Questionnaire* (ASQ) qui comprend douze situations dont la moitié a un devenir favorable, et l'autre moitié un devenir défavorable. Le score de la BDI 21 serait corrélé à l'internalité (0,41), à la stabilité (0,34) et à la globalité (0,35). Mais bien d'autres auteurs ne retrouvent pas de telles corrélations (cité par Eaves et Rush, 1984).

Dans une autre orientation théorique et reprenant les travaux de Sutter sur l'anticipation du déprimé, Légeron, Rivière, Marboutin et Rochat en 1993 ont construit le questionnaire des cognitions anticipatoires (QCA) qui comprend 8 items. Chaque item représente une situation et une cognition associée à partir de laquelle le sujet doit estimer son attitude sur une échelle en quatre points.

Le score global du questionnaire s'étend de 0 à 24. Une valeur seuil de 7 a été définie; un score supérieur à 7 permet de définir avec une grande probabilité un diagnostic d'état dépressif. Envisageons dorénavant quelles sont les utilisations principales de ces questionnaires.

MODES D'UTILISATION DES QUESTIONNAIRES COGNITIFS

Outils diagnostiques

Les scores aux questionnaires servent de critères d'inclusion des patients dans des études expérimentales, permettant d'assurer une certaine homogénéité des populations.

Ainsi, à côté des critères diagnostiques type DSM-III-R, des seuils de sévérité seront définis (par exemple sera inclus tout sujet dont le score à la BDI 21 sera strictement supérieur à 10 points).

Suivre l'évolution des patients sous traitement

De très nombreuses études cliniques se sont succédées pour évaluer l'efficacité des thérapies cognitives. Dobson (1989) répertorie 28 études contrôlées qui concluent de manière homogène à l'efficacité des thérapies cognitives comparées au traitement pharmacologique, le traitement combiné apportant encore de meilleurs résultats. Ces différentes études ont permis de valider le modèle empirique.

Afin d'estimer l'évolution des paramètres cliniques et cognitifs, différents questionnaires et échelles ont été utilisés. Dans la plupart des cas, les hétéro-évaluations font appel aux échelles de Hamilton pour l'anxiété et la dépression. Les auto-évaluations s'effectuent avec l'inventaire pour la dépression de Beck, l'échelle des pensées automatiques, l'échelle des attitudes dysfonctionnelles et quelquefois avec l'échelle de désespoir.

Ces quatre questionnaires cognitifs montrent dans les cas d'amélioration clinique, une diminution de leurs scores totaux, la réduction du score étant plus importante si le sujet déprimé bénéficie d'une thérapie cognitive (Simons, Murphy, Lewin, Wetzler, 1986).

De plus, les retests à distance après six ou douze mois de traitement montrent une fréquence plus élevée de sujets en rémission dans les groupes thérapies cognitives, les critères utilisés étant aussi bien cliniques que métrologiques.

À partir de ces résultats, certains auteurs concluent que la thérapie cognitive entraîne une restructuration des différentes activités cognitives dysfonctionnelles.

Outil thérapeutique

Une autre utilisation des instruments d'évaluation est particulière à la thérapie cognitive. Les questionnaires font partie intégrante du programme thérapeutique ; ils font partie des stratégies de «*feed-back*» qui émaillent tous les entretiens (Mirabel-Sarron et Rivière, 1993). Ainsi, les questionnaires cognitifs (BDI, ATQ, HS et DAS) peuvent être proposés à intervalles réguliers tout au long de la thérapie. Les scores sont donnés aux patients et l'évolution longitudinale du score de chaque item permet d'estimer les changements opérés.

Ils permettent de mettre à jour des difficultés non livrées spontanément par le sujet, certaines cognitions qui, par leur haute fréquence, leur semblaient banales.

Cette information donnée en retour au patient au cours d'entretiens thérapeutiques renforce l'indispensable collaboration patient-thérapeute. Les questionnaires utilisés à titre thérapeutique ne peuvent pas, en même temps, servir d'instruments d'évaluation et obligent le chercheur à faire un choix.

Dépistage des répondeurs et non-répondeurs

Comme dans toute démarche thérapeutique, des facteurs de réponse au traitement ont été recherchés. Aucune des échelles cognitives citées précédemment n'est prédictive de la réponse thérapeutique. Aussi différentes hypothèses cognitives sont-elles alors testées. Mirabel-Sarron, Brun-Eberentz, Samuel-Lajeunesse (1994) utilisent le MMPI pour discriminer les répondeurs avant thérapie et isolent des facteurs spécifiques aux non-répondeurs.

Rosenbaum (1980) émet l'hypothèse que le facteur cognitif d'auto-contrôle issu des théories de Bandura est prédictif de la réponse au traitement ; plus le sujet présente un haut niveau d'auto-contrôle, meilleure est sa réponse thérapeutique.

Aide à la conceptualisation théorique

Le modèle théorique de Beck est toujours l'objet de discussions en dépit de l'efficacité du programme thérapeutique qui en découle. De nos jours, trois tendances existent et s'imbriquent les unes dans les autres.

Le premier mouvement est celui de Beck qui affirme que la théorie cognitive agit bien sur les schémas parce que le taux de rechutes dépressives à distance de la thérapie est bien inférieur à celui constaté avec les médicaments. Simons (1986), Miller (1989), Bowers (1990) observent à six mois un taux de rechute dépressive de 18 %. L'argument principal de Beck est que ce résultat valide sa théorie.

Miranda et Pearson en 1990 attribuent au trouble de l'humeur un rôle interactif. Ils constatent que, chez 47 témoins non déprimés, la DAS est élevée chez ceux qui ont fait dans le passé un épisode dépressif. Chez les autres, le score DAS ne diffère pas de celui des sujets témoins.

Dans cette conception, la DAS ne fournit pas des informations en faveur d'un trait de vulnérabilité, mais reflète une perturbation de l'humeur actuelle ou ancienne.

Un troisième courant est celui des styles attributionnels du déprimé. Abramson, Seligman, Teasdale (1978) démontrent que l'humeur dépressive modifie le style des attributions.

Eaves et Rush (1984) montrent que le style attributionnel distingue significativement les déprimés des témoins. Dans leur étude, le style attributionnel perturbé persiste bien au-delà de la rémission clinique aussi bien chez les déprimés endogènes que non endogènes. Ils supposent l'influence de facteurs de chronicité de la dépression sur le remaniement des styles attributionnels. Étant donné les facteurs multiples (schémas, rôle du trouble de l'humeur, de la chronicité de la dépression), des études sont encore nécessaires pour identifier les variables psychologiques de vulnérabilité à la dépression.

CONCLUSIONS

L'évaluation quantitative des dysfonctionnements cognitifs dans les thérapies cognitives est une voie en développement constant depuis vingt ans aussi bien dans les troubles anxieux que dans les troubles dépressifs; si l'évaluation du contenu des cognitions est de mieux en mieux précisée, c'est l'évaluation des processus et des schémas qui reste à développer et qui permettra un étayage du modèle théorique. Certains auteurs comme Segal et Vella (1990) tentent cette exploration avec des outils propres à la psychologie expérimentale et cognitive comme le test de Stroop.

BIBLIOGRAPHIE

ABRAMSON L. LY., SELIGMAN M. E., TEASDALE J. D. — Learned Helplessness in Humans : Critic and Reformulation,. *Journal of Abnormal Psychology*, *87*, 49-74, 1978.

ALBERT E., LECLERC V., CHNEIWEISS L. — Spécificité des facteurs cognitifs des troubles anxieux aigus et chroniques. *Journal de thérapies comportementales et cognitives*, *4*, 14-21, 1994.

BECK A. T. — *Cognitive Therapy and the Emotional Disorders.* International University Press, New York, 1976.

BECK A. T., BEAMESBERFER A. — Assessment of Depression : the Depression Inventory. *Mod. Prof. in Pharmacopsychiatry*, *7*, 151-159, Karger, Basel, 1974.

BECK A. T., WARD C. H., MENDELSON M., MOCK J. E., ERBAUGH J. K. — An Inventory for Measuring Depression. *Arch. Gen. Psychiatry*, *4*, 561-571, 1961.

BECK A. T., WEISSMAN A., LESTER D., TREIXLER L. — The Measurement of Pessimism : the Hopelessness Scale. *Journal of Consulting and Clinical Psychology*, *42*, 861-865, 1974.

BECK A. T., STEER R., GARBIN M. — Psychometric Properties of the Beck Depression Inventory : 25 Years of Research. *Clinical Psychology Review*, *8 (1)*, 77-100, 1988.

BECK A. T., BROWN G., EIDELSON J., STEER R., RISKIND J. — Differenciating Anxiety and Depression : a Test of the Cognitive Content Specificity Hypothesis. *Journal of Abnormal Psychology*, *96 (3)*, 179-183, 1987.

BECK A. T., STEER R., EPSTEIN N., BROWN G. — Beck Self-Concept Test. *Psychological Assessment*, *2*, 191-197, 1990.

BOBON D., SANCHEZ-BLAQUE A., VON FRENCKELL R. — Comparison between the Short Form of the Beck Depression Inventory and the Combined Beck-Pichot Inventory. *Journal of Psychiatry, Biology and Therapy, 1*, 23-26, 1981.

BOUVARD M., CHARLES S., GUERIN J., AIMARD G., COTTRAUX J. — Étude de l'échelle de désespoir de Beck (hopelessness scale), validation et analyse factorielle. *L'Encéphale, 18*, 237-240, 1992.

BOUVARD M., CHARLES S., MOLLARD E., GUERIN J., AIMARD G., COTTRAUX J. — Étude de validation et analyse factorielle de la version française du questionnaire des pensées automatiques. *Journal de thérapie comportementale et cognitive, 2 (2)*, 25-28, 1992.

BOUVARD M., MOLLARD E. — Version française de l'échelle sociotropie-autonomie de Beck. *Journal de thérapie comportementale et cognitive, 1*, 25-29, 1991.

BOUVARD M., COTTRAUX J., CHARLES S., CIALDELLA P., GUÉRIN J., AIMARD G. — Étude de validation sur une population française de l'échelle d'attitudes dysfonctionnelles de Weissman et Beck (DAS forme A). *Journal de thérapie comportementale et cognitive, 7*, 127-135, 1994.

COLLET L., COTTRAUX J. — Inventaire abrégé de la dépression de Beck (13 items), étude de la validité concurrente avec les échelles de Hamilton et de ralentissement de Widlöcher. *L'Encéphale, 12*, 77-79, 1986.

CRANDELL C. J., CHAMBLESS D. L. — *The Validation of an Inventory for Measuring Depressive Thoughts : the Crandell Chambless Inventory.* Meeting of the Association for Advancement of Behavioral Therapy, Toronto, Canada, 1981.

DOBSON K. S. — A Meta-Analysis of the Efficacy of Cognitive Therapy for Depression. *Journal of Consulting and Clinical Psychology, 57 (3)*, 414-419, 1989.

EAVES G., RUSH A. J. — Cognitive Patterns in Symptomatic and Remitted Unipolar Major Depression. *Journal of Abnormal Psychology*, 1984.

HOLLON S. D., KENDALL P. C. — Cognitive Self-Statements in Depression : Development of an Automatic Thoughts Questionnaire. *Cognitive Therapy and Research, 4*, 383-398, 1980.

KOVACS M., RUSH A. J., BECK A. T., HOLLON S. D. — Depressed Out-Patients Treated with Cognitive Therapy and Pharmacotherapy. A One Year Follow-up. *Arch. Gen. Psychiatry, 38*, 33-39, 1981.

KOVACS M., BECK A. T. — Maladaptative Cognitive Structures in Depression. *Am. J. Psychiatry, 135 (5)*, 525-533, 1978.

LEMPÉRIÈRE T., LÉPINE J. P., ROUILLON F., HARDY P., ADES J., LIUAUTE J. P., FERRAND I. — Comparaison de différents instruments d'évaluation de la dépression à l'occasion d'une étude sur l'Athymil 30. *Ann. médico-psychol., 142*, 1206-1212, 1984.

LEFEBVRE M. F. — Cognition Distortion and Cognitive Errors in Depressed Psychiatric and Lowback Pain Patients. *Journal of Consulting and Clinical Psychology, 49*, 517-525, 1981.

LEGERON P., RIVIÈRE B., MARBOUTIN J. P., ROCHAT C. — Le questionnaire des cognitions anticipatoires : présentation et validation. *L'Encéphale*, XIX, 11-16, 1993.

MILLER I. W., NORMAN W. H., KEITNER G. I. — Cognitive Behavioral Treatment of Depressed Inpatients : Six and Twelve Months Follow-up. *Am. J. Psychiatry, 146*, 1274-1279, 1989.

MIRABEL-SARRON C., RIVIÈRE B. — *Précis de thérapie cognitive.* Dunod, Paris, 1993.

MIRANDA J., PERSONS J. B., BYERS C. N. — Endorsement of Dysfunctional Belief Depends on Current Mood State. *Journal of Abnormal Psychology, 99 (3)*, 237-241, 1990.

PICHOT P., LEMPÉRIÈRE T., MIROUZE — La nosologie des états dépressifs : rapport entre l'étiologie et la sémiologie. Résultats du questionnaire de Beck. *L'Encéphale, 6*, 492-505, 1963.

RUSH A. J., BECK A. T., KOVACS M., WISSENBURGER J., HOLLON S. — Comparison of the Effects of Cognitive Therapy and Pharmacotherapy on Hopelessness and Selfconcepts. *Am. J. Psychiatry*, 862-866, 1982.

ROSENBAUM M. — A Schedule for Assessing Self-Control Behaviors : Preliminary Findings. *Behavior Therapy*, *11*, 104-121, 1980.

SEGAL Z. V., VELLA D. D. — Self-Schema in Major Depression : Replication and Extension of a Priming Methodology. *Cognitive Therapy and Research*, *14 (2)*, 161-176, 1990.

SEGAL Z. V., SWALLOW S. R. — Cognitive Assessment of Unipolar Depression = Measuring Products Processes and Structures. *Behavior Research Therapy*, *32 (1)*, 147-158, 1994.

SILVERMAN J. S., SILVERMAN J. A., EARDLEY D. A. — Do Maladaptive Attitudes Cause Depression? *Arch. Gen. Psychiatry*, *41*, 28-30, 1984.

SIMONS A. D., MURPHY G. E., LEVIN J. L., WETZLER R. D. — Cognitive Therapy and Pharmacotherapy for Depression. Sustained Improvement over One Year. *Arch. Gen. Psychiatry*, *43*, 43-48, 1986.

VERA L., MIRABEL-SARRON C. — *Analyse fonctionnelle et méthodes d'évaluation*. Psychologie Française, sous presse.

WEISSMAN A. N. — *Assessing Depressogenic Attitudes*. Thèse de doctorat de l'université de Pennsylvanie : The Dysfunctional Attitude Scale, 1978.

13

ASPECTS DU VEILLISSEMENT COGNITIF NORMAL

par J. POITRENAUD[1]

Pendant longtemps, les gérontologues ont étudié le *vieillissement cognitif* en le considérant essentiellement comme un phénomène général. Ils se sont alors attachés à en préciser les modalités en comparant les performances cognitives moyennes des sujets âgés avec celles des sujets jeunes. Depuis une douzaine d'années, leur intérêt s'est franchement orienté vers l'analyse des différences interindividuelles existant dans ce phénomène du vieillissement et vers la recherche des facteurs pouvant rendre compte de ces différences (Schaie et Willis, 1986; Svanborg, 1990; Lehr et Thomae, 1991). Cette nouvelle orientation correspond tout à fait à la perspective de prévention dans laquelle se place résolument la gérontologie moderne.

C'est sur ce thème de la *variabilité interindividuelle* qu'il a été choisi de mettre l'accent dans ce chapitre. Cependant, afin de faciliter la compréhension de son contenu, un résumé des recherches consacrées aux aspects généraux du vieillissement cognitif sera tout d'abord présenté.

Cet exposé concernera uniquement le vieillissement normal et sera limité à la période de la vie allant, en gros, de 20 à 80 ans. Le terme « normal » sera utilisé ici, dans son acceptation courante. Il désignera des sujets autonomes, vivant à domicile, ne souffrant pas de maladie somatique invalidante ni d'affection neurologique ou psychiatrique. La limitation à 80 ans est due au fait qu'au-delà de cet âge, il n'existe que très peu d'études systématiques et contrôlées sur le vieillissement cognitif dit « normal ».

Il importe, enfin, de préciser que les recherches dont il sera question ont utilisé des méthodes d'évaluation quantitative reposant sur l'emploi de *tests psychométriques* ou sur les procédures issues des modèles de la psychologie cognitive. Il s'agit donc d'une approche parmi d'autres possibles ; c'est, cependant, celle qui a généralement été utilisée dans l'étude scientifique des modifications cognitives liées à l'âge.

1. Directeur de recherche à l'université de Paris V et responsable de la section de recherche sur le vieillissement différentiel au Centre de gérontologie de l'association Claude Bernard.

ASPECTS GÉNÉRAUX DU VIEILLISSEMENT COGNITIF

Schématiquement, les études réalisées en ce domaine ont eu pour objectifs de répondre aux trois questions suivantes :

1) Existe-t-il réellement un déclin des fonctions cognitives avec l'âge chez le sujet normal?

2) À quel âge se manifestent les premiers signes de ce déclin?

3) Existe-t-il un vieillissement différentiel des fonctions cognitives?

Les réponses apportées à ces questions sont restées longtemps discordantes. Les divergences entre les auteurs étaient essentiellement déterminées par des différences entre les méthodes utilisées pour étudier le vieillissement cognitif. Sans entrer ici dans le détail, nous rappellerons qu'il existe deux grands types d'approche possibles correspondant à deux types d'études : les *études transversales* et les *études longitudinales*.

Une étude transversale consiste à examiner au même moment, par exemple au cours d'une même année, des sujets appartenant à des classes d'âge différentes, c'est-à-dire nés à des époques différentes. Elles reposent sur l'hypothèse que les variations des performances observées entre les classes d'âge reflètent l'effet du vieillissement. Les études transversales sont de loin les plus nombreuses, car relativement faciles à réaliser. Elles ont largement répandu l'idée d'un déclin cognitif pouvant survenir très tôt dans la vie pour certaines aptitudes et ont ainsi contribué à donner du vieillissement une image dévalorisée et dévalorisante.

Une étude longitudinale consiste à suivre les sujets d'une même génération en les examinant à plusieurs reprises au cours de leur vie, les examens successifs étant séparés par des intervalles de plusieurs années. Contrairement aux études transversales, les études longitudinales ne font apparaître que peu de changement dans les fonctions cognitives au cours du vieillissement et suggèrent même que, pour certaines fonctions, les performances peuvent s'améliorer jusqu'à un âge relativement avancé. De tels résultats ont amené certains auteurs à conclure que le déclin intellectuel avec l'âge était un mythe.

L'approche longitudinale est, certes, la meilleure pour étudier les changements liés à l'âge mais elle n'est pas sans soulever certaines critiques. En fait, l'approche transversale et l'approche longitudinale comportent l'une et l'autre des avantages et des inconvénients qui permettent de dire que, d'une façon générale, les premières surestiment l'effet de l'âge alors que les secondes le sous-estiment.

Certains auteurs, comme Schaie et ses collaborateurs, ont élaboré des stratégies complexes[2], combinant l'approche transversale et l'approche longitudinale afin d'éliminer les inconvénients méthodologiques. Malheureusement, aucune de ces stratégies ne constitue une panacée, et il n'existe pas, actuellement, de méthode parfaite pour étudier le vieillissement cognitif.

Cependant, les points de vue des chercheurs se sont rapprochés et l'on considère maintenant que la vérité se situe probablement entre les conclusions des études transversales et celles des études longitudinales.

2. Ces stratégies sont décrites dans le chapitre 13 de J. POITRENAUD (1987) cité plus loin.

À partir d'une analyse de la littérature récente, on peut proposer le bilan suivant concernant les réponses aux trois premières questions :

1) Le déclin cognitif n'est pas un mythe, mais une réalité.

2) Ce déclin est moins précoce et son amplitude est moins importante que ne l'indiquaient les conclusions des premières recherches.

3) Toutes les fonctions cognitives ne se modifient pas de la même manière au cours du vieillissement.

Le tableau 1 résume et illustre ces informations. Il indique, pour chacune des fonctions, l'âge moyen auquel se manifestent cliniquement les premiers signes d'un déclin.

Tableau 1. *Évolution différentielle des aptitudes cognitives (d'après une synthèse personnelle des données de la littérature)*

Aptitudes	*Âge moyen auquel se manifestent les premiers signes d'un déclin*			
	< 55	55-64	65-74	≥ 75
• Domaine intellectuel				
compréhension verbale				×
fluidité verbale			×	
raisonnement arithmétique				×
raisonnement abstrait		×	×	
aptitudes visuo-spatiales				
• Mémoire				
tertiaire				×
secondaire épisodique	×			
primaire			×	
• Attention				
tâches simples ou automatiques			×	
tâches complexes nouvelles	×			
• Rapidité psychomotrice		×		

On voit que dans le domaine intellectuel, il ne se passe à peu près rien jusqu'à la soixantaine. Certaines fonctions, comme la compréhension verbale et le raisonnement arithmétique, se maintiennent jusqu'à un âge déjà avancé (75 ans et plus). Il en est de même de la *mémoire tertiaire* (c'est-à-dire la mémoire des faits anciens, et certains aspects de la mémoire sémantique) et pour l'attention lorsqu'il s'agit de tâches simples et automatisées. En revanche, des fonctions telles que la *mémoire épisodique secondaire* (c'est-à-dire la mémoire des faits et des événements vécus récents), l'attention lorsqu'il s'agit de tâches complexes et (ou) nouvelles, et la rapidité d'intégration psychomotrice apparaissent beaucoup plus fragiles. Pour certaines d'entre elles, le déclin se manifeste avant 55 ans et même peut-être plus tôt selon certains auteurs. La mémoire primaire, c'est-à-dire la *mémoire immédiate*, occupe une position intermédiaire.

Il est important de noter que la mémoire épisodique secondaire, l'attention dans les tâches complexes et l'intégration psychomotrice sont des fonctions en étroite interaction. Il est difficile de savoir laquelle joue le rôle de « starter » dans le déclin. Plusieurs hypothèses ont été proposées à ce sujet. Pour certains, la cause principale de ces modifications cognitives réside dans la diminution de la rapidité mentale avec l'âge ; pour d'autres, le facteur essentiel est la diminution des ressources attentionnelles ; pour d'autres enfin, il s'agit surtout d'un déficit dans les stratégies mises en œuvre pour acquérir des informations nouvelles. En fait, il est sans doute préférable de retenir la notion d'interaction (ou de causalité réciproque) et de considérer que c'est le complexe attention-mémoire-rapidité psychomotrice qui apparaît comme particulièrement sensible aux effets de l'âge.

Ce tableau appelle toutefois certains commentaires :

1) Les informations qu'il contient proviennent d'une synthèse personnelle des résultats de plusieurs recherches longitudinales ou séquentielles réalisées par différentes équipes gérontologiques. Aucune étude, en effet, n'a appréhendé l'ensemble des modifications cognitives pour la période allant de 20 à 80 ans. Ces informations ont une valeur seulement indicative et ne doivent pas être prises au pied de la lettre.

2) La configuration représentée par l'évolution différentielle des aptitudes a, par contre, de bonnes chances d'être très proche de la réalité. Tous les auteurs s'accordent pour dire que le complexe mémoire-attention-rapidité mentale constitue le « noyau » très sensible aux effets de l'âge.

3) Ce tableau concerne ce qu'on observe, en moyenne dans les modifications cognitives. Or, il existe d'importantes différences individuelles en ce domaine.

DIFFÉRENCES INTERINDIVIDUELLES

Dans son étude longitudinale où il suit certains de ses sujets depuis 35 ans, Schaie (1990) a, par exemple, constaté que 25 % environ des septuagénaires avaient réalisé les mêmes performances psychométriques à un ensemble de cinq épreuves intellectuelles que 7 ans auparavant. C'était encore le cas de 12 à 15 % d'octogénaires. D'autres sujets avaient, en revanche, décliné de façon plus ou moins globale ; certains (mais il s'agissait d'une petite minorité (2 à 3 %)) avaient décliné à tous les tests. Des constatations semblables se sont dégagées des deux études longitudinales que j'ai réalisées avec mon équipe.

Comme je l'ai déjà mentionné, l'objectif important des études gérontologiques est l'identification des facteurs (médicaux, psychologiques, sociaux ou autres) susceptibles d'expliquer ces différences interindividuelles. Une analyse de la littérature, réalisée il y a quelques années (Poitrenaud, 1987), m'avait permis de retenir cinq facteurs comme déterminants possibles de cette variabilité interindividuelle. Deux d'entre eux, l'existence d'une maladie cardio-artérielle et une consommation quotidienne abusive d'alcool semblaient avoir une influence néfaste. Trois autres, en revanche, paraissaient avoir une influence positive : un niveau élevé ou assez élevé d'activité physique ; un niveau élevé de stimulation cognitive ; un niveau d'études supérieur ou très supérieur à la moyenne.

ÉTUDE PERSONNELLE

Il y a environ dix-huit ans, j'ai entrepris avec mes collaborateurs, une étude longitudinale, dans le but de vérifier l'influence de tels facteurs et de quelques autres. Les résultats de cette étude sont les suivants.

Sujets

L'étude a porté sur un groupe de sujets de sexe masculin appartenant à une catégorie socioprofessionnelle homogène, constituée de cadres moyens et supérieurs de l'industrie et du commerce. Ce groupe a été examiné à trois reprises entre 1976 et 1990. Chaque examen comprenait des tests psychométriques destinés à évaluer certains aspects du fonctionnement cognitif, ainsi que diverses investigations médicales et psychosociales permettant d'évaluer l'état de santé des participants, leur style de vie et diverses caractéristiques de leur environnement sociofamilial.

Les critères de sélection de la population d'étude ont été les suivants : les participants devaient être de sexe masculin, âgés de 63 ou 64 ans, travailler à temps complet, avoir un salaire mensuel égal ou supérieur à 7000 F en 1975 (environ 25000 F en 1994), travailler à Paris, résider dans l'agglomération parisienne. Pour recruter ces sujets, nous nous sommes adressés à la Caisse interprofessionnelle de prévoyance des cadres, la principale caisse de cadres française.

Deux cent vingt sujets de cette caisse répondaient à nos critères de sélection; 180 d'entre eux (83 %) acceptèrent de participer à l'étude et vinrent dans notre laboratoire pour un examen médico-psychologique. Sept ans plus tard (T7), 125 de ces 180 volontaires purent être examinés une seconde fois et 82 d'entre eux, une troisième fois, 14 ans après le début de l'étude (T14). Cet amenuisement de l'effectif, phénomène classique dans les études longitudinales, a été dû à des causes diverses : déménagement hors Paris, maladie, refus de continuer à participer, décès.

Sur les 82 participants restants, 7 avaient souffert d'un accident vasculaire cérébral durant le déroulement de l'étude. Étant donné que ce type d'accident peut entraîner une altération pathologique du fonctionnement cognitif, ces sept sujets ont été exclus de la présente recherche. Deux autres ont été également exclus, l'un parce qu'il ne put réaliser les tests psychométriques en raison d'une détérioration mentale associée à une maladie de Parkinson, l'autre parce qu'il refusa de passer l'examen psychométrique à T14. Pour toutes ces raisons, notre étude a finalement porté sur un échantillon de 73 sujets.

Mesures

Tests cognitifs

Cinq tests psychométriques classiques, destinés à évaluer la mémoire, la rapidité d'intégration psychomotrice et certains aspects de l'attention ont été sélectionnés. Il s'agit des épreuves suivantes :

a) Le test de rétention visuelle de Benton (TRV; administration A; système de cotation de Poitrenaud et Moreaux). La forme C du test a été utilisée à T0 et T7, la forme D à T14 afin de minimiser les effets d'apprentissage.

b) Une version légèrement modifiée (trois essais au lieu de cinq) du test de mémorisation d'une série de 15 mots (15M) de Rey. La forme A a été utilisée à T0 et T7, la forme C à T14. À chaque temps, la note au troisième essai a été le paramètre retenu.

c) Le test de code de Clément (CODE), la forme C ayant été administrée à T0 et T7, la forme D, à T14.

d) Le *Trail Making Test*. Les parties A (TMA) et B (TMB) ont été utilisées comme deux épreuves séparées dans cette étude. Pour les deux parties, la note a été le temps (en secondes) mis pour accomplir la tâche. Afin de normaliser la distribution des notes, les scores bruts ont été convertis en scores standardisés selon une échelle allant de 1 à 10 (10 correspondant à la meilleure performance, c'est-à-dire au temps le plus court). Le *Trail Making Test* ne comportant pas de formes parallèles, les mêmes formes ont été utilisées pour TMA et TMB aux trois temps de l'étude.

Déterminants possibles des différences interindividuelles dans le vieillissement cognitif

Les six facteurs suivants ont été sélectionnés et évalués comme indiqué ci-dessous :

a) Existence d'une maladie cardio-artérielle. Un sujet était considéré comme souffrant d'une affection cardio-artérielle si au moins un des critères suivants était présent : une pression artérielle systolique égale ou supérieure à 160 mmHg ou une pression diastolique égale ou supérieure à 100 mmHg; une maladie cardio-artérielle avérée, constatée le jour de l'examen ou mentionnée dans les antécédents personnels, telle que : infarctus du myocarde, maladie coronarienne, artérite, etc.

Ce facteur « maladie cardio-artérielle » a été évalué à chacun des trois temps de l'étude (T0, T7, T14) selon l'échelle dichotomique suivante : 0 : pas de maladie cardio-artérielle; 1 : existence d'une telle maladie (actuellement ou dans les antécédents).

b) Le niveau d'éducation scolaire, évalué à T0 sur une échelle en trois points : 0 : moins de huit ans de scolarité; 1 : de 8 à 12 ans; 3 : treize ans ou plus.

c) L'activité physique, mesurée, à T7 et T14, par le nombre d'heures consacrées chaque semaine à des activités impliquant un exercice physique, telles que : marche, jardinage, bricolage, travaux domestiques, gymnastique ou sports.

d) L'activité socioculturelle, mesurée, à T7 et T14, par le nombre d'heures passées chaque semaine à des activités telles que : lecture, jeux impliquant une gymnastique mentale, activité professionnelle de type intellectuel, activités artistiques ou participation à l'organisation et au fonctionnement d'associations sociales ou religieuses.

e) La consommation d'alcool, mesurée par la quantité d'alcool (en grammes) absorbée quotidiennement (d'après les dires du sujet). Ce facteur n'a été évalué qu'à T14.

f) L'usage de somnifères, évalué à T7 et T14 par la fréquence de la consommation, selon l'échelle d'appréciation suivante : 1 : jamais ou moins d'une fois par mois ; 2 : de une à trois fois par mois ; 3 : une fois par semaine ; 4 : plusieurs fois par semaine ; 5 : quotidiennement.

Analyse des données

Pour chacun des tests cognitifs, les performances moyennes à T0, T7 et T14 ont été comparées par une analyse de variance avec mesures répétées, incluant le test de Newman-Keuls.

L'analyse des relations entre les changements cognitifs et leurs possibles déterminants a soulevé un problème méthodologique : lequel des trois temps de mesure (T0, T7 ou T14) pouvait être considéré comme le meilleur pour obtenir l'évaluation la plus satisfaisante de l'influence de ces déterminants? Comme le processus de vieillissement cognitif peut être déterminé aussi bien par des facteurs du passé que par des facteurs actuels, et comme la plupart des facteurs que nous avions retenus comme déterminants possibles n'avaient pas été évalués à T0, il fut décidé de réaliser séparément deux séries d'analyses : la première basée sur la mesure des déterminants à T7, et la seconde sur leur mesure à T14. Pour chaque série, ce sont les relations entre les déterminants et les changements survenus dans les performances psychométriques entre T0 et T14 qui ont été analysées.

Dans les deux cas, la méthode statistique utilisée pour évaluer ces relations a été la méthode d'analyse du score de changement résidualisé. La méthode a consisté à effectuer, pour chacun des tests cognitifs, une analyse de régression multiple où le score à T14 était la variable dépendante et le score à T0, la première variable indépendante entrée dans l'équation. Cette méthode a permis d'évaluer de façon indépendante le lien existant entre chacun des déterminants possibles et les changements cognitifs, et aussi de classer hiérarchiquement les déterminants selon l'intensité de ce lien.

Résultats

Les résultats de l'étude seront présentés ici de façon résumée, afin de respecter la place impartie à chacun des auteurs de cet ouvrage. Un exposé complet des données de notre recherche peut être trouvé dans un article récemment publié (Poitrenaud *et al.*, 1994).

Comparaison des performances psychométriques moyennes à T0, T7, T14

Pour chaque test, les moyennes et les écarts-types des notes à chaque point de mesure sont donnés dans le tableau 2 avec les résultats de l'analyse de variance avec mesures répétées.

Tableau 2. *Moyennes et écarts-types des notes aux tests psychométriques à chaque temps de mesure, avec résultats de l'analyse de variance avec mesures répétées (N = 73)*

	Temps de mesure						
	T0		*T7*		*T14*		
Test	M	σ	M	σ	M	σ	F (ANOVA)
TRV	20,52	(2,42)	20,68	(2,11)	18,15	(3,66)	33,85***
15M	11,45	(2,06)	10,49	(2,44)	8,95	(2,30)	52,49***
Code	66,19	(12,03)	64,90	(11,85)	59,00	(14,54)	18,84***
TMA	7,40	(1,67)	7,84	(1,65)	6,90	(2,20)	7,97***
TMB	7,42	(2,01)	7,59	(2,97)	6,84	(2,11)	7,08***

*** $p < 0,001$

Notes : TRV = test de rétention visuelle de Benton; 15M = test de mémorisation de 15 mots de Rey; Code = test de code de Clément; TMA = *Trail Making Test*, partie A; TMB = *Trail Making Test*, partie B.

De ce tableau, il ressort que :

a) Pour tous les tests, le score moyen a significativement décliné entre T0 et T14.

b) Toutefois, ce déclin n'est pas linéaire : il n'existe pratiquement pas de changement entre T0 et T7, sauf pour le test de mémorisation verbale (15M) pour lequel on constate une baisse légère mais significative ($p < 0,01$). En revanche, pour tous les tests, la performance moyenne a décliné très significativement entre T7 et T14, c'est-à-dire entre 71 et 78 ans. La baisse la plus marquée est observée pour les tests de mémorisation.

c) La légère augmentation (non significative) de la performance entre T0 et T7, constatée pour deux des tests (TMA et TMB) s'explique probablement par un effet d'apprentissage.

Cependant, les résultats précédents concernent seulement les performances moyennes. Il existe, en fait, d'importantes différences interindividuelles dans les modifications cognitives survenues entre T0 et T14. Pour le TRV, par exemple, 22 des 73 participants ont obtenu à T14 une note égale ou supérieure à celle obtenue à T0, tandis que 18 autres ont perdu 5 points ou plus entre les deux points de mesure. Une variabilité de même amplitude a été observée pour les quatre autres tests.

Relations entre les modifications psychométriques liés à l'âge et leurs possibles déterminants

Les deux séries d'analyses que nous avons effectuées ont fourni des résultats pratiquement identiques. Pour simplifier, nous ne présenterons donc ici que les résultats portant sur l'évaluation des déterminants à T14. Notons, cependant que, dans la mesure où les deux séries d'analyses étaient indépendantes, cette similitude tend à confirmer la validité des données recueillies.

Tableau 3. *Relations entre les modifications psychométriques liées à l'âge et leurs déterminants possibles : valeurs des coefficients Bêta[1] significatifs. Analyses basées sur la mesure des déterminants à T14 (N = 73)*

	Tests psychométriques				
Déterminant[2]	TRV	15M	Code	TMA	TMB
Maladie cardio-artérielle	–0,27**			–0,25*	–0,20*
Niveau d'études	0,23*				0,18*
Activité physique		0,22*		0,23*	0,36***
Activité socioculturelle					
Somnifères					
Alcool					

* p < 0,05 ** p < 0,01 *** p < 0,001

[1] Coefficients de régression standardisés.
[2] Un coefficient Bêta négatif indique un effet négatif.

Notes : TRV = test de rétention visuelle de Benton; 15M = test de mémorisation de 15 mots de Rey; Code = test de code de Clément; TMA = *Trail Making Test*, partie A; TMB = *Trail Making Test*, partie B.

Les résultats de l'analyse du score de changement résidualisé sont résumés dans le tableau 3. Celui-ci donne la valeur des coefficients de régression standardisés « Bêta » qui sont apparus statistiquement significatifs. Les coefficients « Bêta » expriment le lien existant entre les déterminants et les modifications psychométriques observées entre T0 et T14. Un coefficient significatif indique que le déterminant explique un pourcentage significatif de la variance des changements cognitifs. Un coefficient positif suggère un effet favorable du déterminant sur l'évolution des performances, un coefficient négatif suggère un effet défavorable.

Les données du tableau 3 montrent que trois des déterminants que nous avions sélectionnés sont significativement liés à l'évolution des performances psychométriques entre T0 et T14. Le facteur maladie cardio-artérielle est lié négativement aux changements cognitifs observés au TRV ($p < 0,01$), au TMA ($p < 0,05$) et au TMB ($p < 0,05$). Le niveau d'éducation scolaire est positivement lié aux modifications des performances au TRV et au TMB ($p < 0,05$ dans les deux cas). L'activité physique est également associée positivement aux changements des performances aux tests des 15M ($p < 0,05$), au TMA ($p < 0,05$) et au TMB ($p < 0,001$). Il est intéressant de souligner que pour le TMB, l'activité physique explique une part substantielle de la variance des changements observés (13,4 %).

Aucun des trois autres déterminants possibles (activité socioculturelle, consommation quotidienne d'alcool et consommation de somnifères) n'est apparu statistiquement corrélé avec les changements psychométriques survenus entre T0 et T14.

Discussion

Avant de commenter les résultats qui viennent d'être brièvement exposés, quelques remarques s'imposent concernant les caractéristiques de notre recherche.

Il convient tout d'abord d'en souligner les limites.

L'échantillon étudié (composé de sujets de sexe masculin, cadres parisiens de l'industrie et du commerce, travaillant encore à temps complet au début de l'étude) n'était pas représentatif de la population générale âgée. En outre, comme dans toute recherche longitudinale, un nombre important de sujets est sorti de l'étude avant son terme. Les analyses ont donc été effectuées sur un groupe sélectionné qui n'était sans doute que partiellement représentatif de l'échantillon initial.

L'investigation n'a porté que sur certains aspects du vieillissement cognitif. La batterie de tests utilisée était censée évaluer la mémoire épisodique secondaire, l'attention dans les tâches complexes et l'intégration psychomotrice, toutes fonctions très sensibles aux effets de l'âge. Cependant, d'autres fonctions, telles que, par exemple, le raisonnement verbal ou non verbal, la compréhension et la fluidité verbales ou encore l'organisation visuo-spatiale, n'ont pas été explorées.

La période relativement courte sur laquelle les changements cognitifs liés à l'âge ont été évalués (en gros de 63 à 78 ans) constitue une autre limite de cette étude.

En revanche, cette étude présente certains avantages. En premier lieu, elle porte sur un échantillon très homogène en ce qui concerne le statut socio-économique et l'âge. On sait que le statut socio-économique est probablement un important déterminant du vieillissement cognitif différentiel (Lehr et Thomae, 1991) pouvant masquer l'influence d'autres facteurs. Le fait qu'il ait été, ici, contrôlé (tous les participants appartenant à une catégorie socioprofessionnelle supérieure) a permis de mieux étudier l'influence d'autres déterminants. Similairement, comme tous les participants avaient le même âge, il a été possible d'évaluer les changements liés à l'âge, indépendamment de la plupart des effets de cohorte. Un autre avantage a été l'emploi d'une méthode statistique permettant d'apprécier le rôle de chaque déterminant indépendamment de l'influence possible des autres déterminants.

Toutes ces considérations doivent être présentes à l'esprit dans la discussion des résultats obtenus.

Changements liés à l'âge dans les performances cognitives

Le score moyen a décliné significativement pour tous les tests entre T0 et T14. Cependant, il est apparu que ce déclin n'était devenu substantiel que vers le milieu de la huitième décennie. Si l'on se réfère aux données recueillies par d'autres auteurs (Arenberg et Robertson-Tchabo, 1977 ; Schaie et Hertzog, 1983) avec le même type d'instruments de mesure, la baisse des performances psychométriques avec l'âge semble survenir plus tôt dans la population générale. Cette constatation appuie l'hypothèse selon laquelle les individus appartenant aux catégories socioculturelles supérieures maintiennent, en général, plus longtemps que les autres, un haut niveau de fonctionnement cognitif.

Cependant, nos résultats ont également mis en évidence qu'il existait d'importantes différences interindividuelles dans les modifications psychométriques liées à l'âge chez les participants à notre étude. Le but principal de cette recherche longitudinale était, rappelons-le, de contribuer à l'identification des facteurs susceptibles de rendre compte de cette variabilité.

Relations entre les changements cognitifs et leurs déterminants possibles

Trois facteurs parmi ceux que nous avions sélectionnés comme déterminants possibles sont apparus significativement liés aux modifications psychométriques observées entre T0 et T14. La question importante posée par ces résultats est de savoir quelle est la nature de ces relations. Le caractère corrélationnel de notre étude ne permet pas de conclure à l'existence d'un effet causal direct du niveau d'éducation scolaire, des maladies cardio-artérielles et de l'activité physique sur les modifications cognitives liées à l'âge.

Concernant le niveau d'éducation scolaire, cet effet paraît toutefois vraisemblable, et nos résultats suggèrent fortement qu'un niveau élevé favorise le maintien de certaines aptitudes cognitives. Un tel effet est, toutefois, quelque peu surprenant dans un groupe de sujets âgés appartenant tous à une catégorie socioprofessionnelle supérieure. Il semble souligner l'importance des connaissances et des capacités acquises pendant les années de scolarité pour la conservation, de nombreuses années plus tard, d'un haut niveau de performance dans certaines tâches cognitives. Il est, en outre, intéressant de noter que cette influence apparemment favorable du niveau d'éducation scolaire a été également observée par Rudinger, cité par Lehr et Thomae (1991), dans une étude longitudinale réalisée sur un échantillon représentatif de la population générale de l'Allemagne de l'Ouest.

Pour les maladies cardio-artérielles, l'hypothèse d'une causalité directe est plus discutable. Même si l'on fait abstraction du rôle possible de certains médicaments (notamment antihypertenseurs) dans les modifications cognitives observées, d'autres interprétations peuvent être proposées. Certains auteurs (Hertzog, Schaie et Gribbin, 1978) suggèrent que l'influence des maladies cardio-artérielles sur le fonctionnement cognitif s'exerce peut-être de façon indirecte par l'intermédiaire des modifications qu'elles entraînent dans le genre de vie des patients, en réduisant par exemple leurs activités. Notons, cependant, que nos données n'appuient pas cette interprétation dans la mesure où l'influence des maladies cardio-artérielles a été évaluée indépendamment de celle des activités physiques et socioculturelles. Quoi qu'il en soit, nos résultats concernant les relations entre maladies cardio-artérielles et modifications cognitives liées à l'âge, sont en accord avec ceux de l'étude longitudinale de Hertzog *et al.* (1978) pour qui l'existence d'une affection cardio-vasculaire pouvait être considérée comme un facteur de risque du déclin cognitif.

Concernant l'activité physique, plusieurs études expérimentales (Clarkson-Smith et Hartley, 1989; Shay et Roth, 1992) ont déjà suggéré qu'une activité physique régulière pouvait avoir des effets bénéfiques sur les performances des sujets âgés à certaines tâches cognitives. Si nos données vont dans le sens de cette hypothèse, elles ne permettent pas, bien entendu, de conclure à un effet causal direct de la fréquence de l'activité physique sur les changements cognitifs associés à l'âge. Il est, en effet, possible que les relations observées entre l'activité physique et les modifications psychométriques s'expliquent par leur liaison commune avec d'autres facteurs, non étudiés ici, comme, par exemple, certains traits de personnalité.

Aucun des trois autres déterminants possibles (activité socioculturelle, consommation d'alcool et usage de somnifères) n'a été trouvé significativement lié aux modifications psychométriques survenues entre T0 et T14. Ce résultat ne signifie pas nécessairement que ces trois facteurs ne jouent aucun rôle dans les changements cognitifs associés à l'âge.

Les résultats de certaines recherches (Gribbin, Schaie et Parham, 1980; Schaie, 1984) ayant montré que la poursuite d'activités intellectuelles au cours du vieillissement était positivement corrélée avec le maintien d'un haut niveau de fonctionnement cognitif, nous avions formulé l'hypothèse d'une liaison positive entre une fréquence élevée d'activité socioculturelle et le maintien d'un tel fonctionnement. Le fait que cette hypothèse n'ait pas été confirmée dans notre recherche s'explique peut-être par les caractéristiques de notre population d'étude. En effet, le nombre d'heures consacrées par semaine à des activités socioculturelles par les 73 participants était, en moyenne, de 24 heures et, aussi bien à T7 qu'à T14, seuls quatre d'entre eux consacraient moins de 7 heures par semaine à ces activités. L'effet néfaste possible d'un bas niveau d'activité socioculturelle sur le maintien des fonctions cognitives avec l'âge avait donc peu de chance d'être mis en évidence dans cet échantillon.

Les données de plusieurs études (Ryan, 1982; Poitrenaud, Vallery-Masson et Burnat, 1986) ont suggéré que l'alcoolisme chronique pouvait entraîner un vieillissement prématuré des fonctions cognitives chez les sujets âgés. Dans une recherche (Poitrenaud *et al.*, 1986) portant sur 313 sujets de sexe masculin, âgés de 60 à 63 ans, nous avions observé une corrélation significative et négative entre les performances réalisées à plusieurs tests cognitifs et l'absorption quotidienne d'alcool. C'est pourquoi nous avions formulé l'hypothèse d'une corrélation négative et significative entre ce facteur et le maintien de certaines fonctions cognitives chez les participants à la présente étude. Là encore, le fait que cette hypothèse n'ait pas été vérifiée peut s'expliquer par les caractéristiques de l'échantillon étudié. En effet, la consommation quotidienne d'alcool était, en moyenne, modérée (environ 36 g) chez les 73 participants, 82 % d'entre eux consommant moins de 70 g.

La grande majorité des sujets qui, dans cette étude, prenaient des somnifères, utilisaient des benzodiazépines. Les résultats de certains travaux (Closser, 1991) ayant montré que la dépendance et les perturbations cognitives et psychomotrices induites par ces médicaments étaient plus importantes chez les sujets âgés que chez les jeunes, nous nous attendions à observer une relation négative entre la fréquence de la consommation de somnifères et le maintien des performances à certains tests psychométriques. En fait, nous avons constaté que la consommation de somnifères était bien liée négativement au changement de la performance au test de Code et au TMB, mais ces liaisons n'étaient pas statistiquement significatives.

Pour clore cette discussion, il importe de souligner le caractère relatif des résultats fournis par notre étude longitudinale. L'examen du tableau 3 montre que si les corrélations observées entre, d'une part, le niveau d'éducation scolaire, les maladies cardio-artérielles, l'activité physique et, d'autre part, les changements psychométriques, étaient toutes significatives, leur valeur était relativement faible. Cela signifie, en d'autres termes, que ces trois facteurs n'expliquaient qu'une part relativement faible de la variance de ces changements. Même lorsqu'ils étaient liés significativement tous les trois au changement de la performance à un même test (en l'occurrence le TMB), le pourcentage de variance expliquée n'excédait pas 30 %. Une partie substantielle des changements psychométriques restait donc inexpliquée. De nouvelles études sont nécessaires pour tenter d'identifier d'autres facteurs (génétiques, nutritionnels ou de personnalité, par exemple) susceptibles de rendre compte des différences interindividuelles dans le vieillissement cognitif.

Bibliographie

ARENBERG D., ROBERTSON-TCHABO E. A. — Learning and Aging. In : BIRREN J. E., SCHAIE K. W. (Eds). *Handbook of the Psychology of Aging*. Van Nostrand Reinhold, New York, 421-449, 1977.

CLARKSON-SMITH L., HARTLEY A. A. — Relationships between Physical Exercise and Cognitive Abilities in Older Adults. *Psychology and Aging*, *4*, 183-189, 1989.

CLOSSER M. H. — Benzodiazepines and the Elderly. A Review of Potential Problems. *Journal of Substance Abuse Treatment*, *8 (1-2)*, 35-41, 1991.

GRIBBIN K., SCHAIE K. W., PARHAM I. A. — Complexity of Life Style and Maintenance of Intellectual Abilities. *Journal of Social Issues*, *21*, 47-61, 1980.

HERTZOG C., SCHAIE K. W., GRIBBIN K. — Cardiovascular Disease and Changes in Intellectual Functioning from Middle to Old Age. *Journal of Gerontology*, *33*, 872-883, 1978.

LEHR U., THOMAE H. — Aging in Europe : New Directions in Psychology. *European Journal of Gerontology*, *1*, 43-52, 1991.

POITRENAUD J. — Les modifications psychométriques au cours du vieillissement cérébral normal (problèmes méthodologiques et résultats). In : GUEZ D., *Investigations cliniques et paracliniques dans le vieillissement cérébral*, Doin, Paris, 17-34, 1987.

POITRENAUD J., VALLERY-MASSON J., BURNAT G. — Consommation d'alcool et performances cognitives chez des sujets à l'âge de la retraite. *Bulletin de la société française d'alcoologie*, *8*, 27-32, 1986.

POITRENAUD J., VALLERY-MASSON J., DARCET P., BARRERE H., DERRIENNIC F., GUEZ D. — Sources of Individual Differences in Cognitive Aging : a Longitudinal Study of an Elderly French Managerial Population. In : VELLAS B. J., ALBAREDE J. L., GARRY P. J., *Facts and Research in Gerontology*, Serdi Publisher, Paris, 35-50, 1994.

RYAN C. — Alcoholism and Premature Aging : a Neuropsychological Perspective. *Alcoholism : Clinical and Experimental Research*, *6*, 22-30, 1982.

SCHAIE K. W. — Midlife Influences upon Intellectual Functioning in Old Age. *International Journal of Behavioral Development*, *7*, 463-478, 1984.

SCHAIE K. W. — Intellectual Development in Adulthood. In : BIRREN J. E., SCHAIE K. W., *Handbook of the Psychology of Aging*, Academic Press, San Diego, 291-309, 1990.

SCHAIE K. W., HERTZOG C. — Fourteen Year Cohort-Sequential Analyses of Adult Intellectual Development. *Developmental Psychology*, *19*, 531-543, 1983.

SCHAIE K. W., WILLIS S. L. — Can Decline in Adult Intellectual Functioning be reversed? *Developmental Psychology*, *22*, 223-232, 1986.

SHAY K. A., ROTH D. L. — Association between Aerobic Fitness and Visuo-spatial Performance in Healthy Older Adults. *Psychology and Aging*, *7*, 15-24, 1992.

SVANBORG A. — Longitudinal Aspects of Aging. *Clinical and Scientific Psychogeriatrics*, *1*, 112-116, 1990.

14

ANALYSE DES DONNÉES EN PSYCHOPATHOLOGIE QUANTITATIVE

par P. BOYER

Le problème de l'analyse des données en psychopathologie quantitative est source de controverses fréquentes entre cliniciens et statisticiens. Il est souvent reproché aux cliniciens de ne pas formuler clairement les hypothèses qui doivent être testées et d'outrepasser leurs droits en matière d'interprétation des résultats.

Afin de rendre plus concret ce débat et de passer en revue les différents pièges qu'un clinicien peut trouver sur son chemin lorsqu'il veut analyser ses résultats cliniques, nous prendrons l'exemple d'un corpus de données précis, celui d'une étude épidémiologique menée par l'OMS en médecine générale.

Nous ne fournirons dans ce chapitre aucun résultat concernant la prévalence des troubles et les différents scores aux échelles utilisées dans cette étude car une discussion de nature épidémiologique n'est pas notre objet. L'enquête OMS nous servira ici seulement de support à « l'exemplification » (si ce néologisme est autorisé !) de l'analyse de données multiples en psychopathologie quantitative.

LES DONNÉES

Résultats du centre français participant à l'étude internationale sur l'évaluation diagnostique et thérapeutique des troubles psychologiques chez des patients consultant en médecine générale.

Buts de l'étude

Le but de cette étude internationale coordonnée par l'OMS était :

– de déterminer la prévalence des principaux troubles mentaux non psychotiques dans une population de sujets consultant en médecine générale ;

– de déterminer la part relative des troubles majeurs (caractérisés, au-dessus du seuil de diagnostic) et mineurs (au-dessous du seuil de diagnostic) ;

– d'analyser la reconnaissance et le traitement de ces troubles par le médecin généraliste ;

– de déterminer quel a été le type d'accès aux soins des consultants ;

– de comparer sur ces critères les 14 centres internationaux ayant participé à l'étude.

Déroulement de l'étude

Cette étude collaborative s'est déroulée en trois étapes :

– *1re étape* : passation du questionnaire de santé générale de D. Goldberg (GHQ) par des consultants consécutifs au cabinet du praticien généraliste (ce questionnaire comporte 12 items correspondant à des plaintes de nature psychopathologique); recueil des données socio-démographiques; parallèlement, passation par le médecin généraliste d'un certain nombre de fiches diagnostiques et de traitement.

– *2e étape* : après procédure d'échantillonnage, passation d'un entretien diagnostique standardisé, le CIDI (OMS, 1992) permettant dans la version de cette étude de faire, grâce à des algorithmes, les principaux diagnostics non psychotiques selon l'ICD-10 et le DSM-III-R. Lors de la même étape, passation de questionnaires concernant le handicap social.

– *3e étape* : suivi à 3 mois et à 12 mois des patients ayant passé l'entretien.

Deux psychologues et trois psychiatres ont conduit la formation de 12 assistants de recherche qui ont assisté le médecin généraliste dans la première étape et qui ont réalisé l'entretien standardisé de la seconde étape. Les médecins généralistes ont été réunis au début de l'étude pour information sur les contraintes de l'enquête, puis plusieurs fois au cours de l'étude.

Sélection de l'échantillon

2699 patients ont été approchés pour l'étude en France. Parmi eux, 548 patients ne remplissaient pas les critères d'inclusion. 2151 patients ont donc été retenus. Et 55 d'entre eux ont refusé de passer le questionnaire de santé générale de Goldberg (GHQ-12). 2096 GHQ-12 ont finalement été recueillis. Les scores au GHQ des patients de cette première étape se répartissent comme suit : 61,9 % de l'échantillon ont un score bas (0-3), 22,0 % un score moyen (4-6) et 16,1 % un score élevé (7 à 12).

La procédure d'échantillonnage à partir des scores obtenus au GHQ a identifié 654 patients éligibles pour un entretien diagnostique standardisé de seconde étape. 38,1 % des patients éligibles ayant refusé ou ayant été dans l'impossibilité d'avoir un entretien approfondi, l'échantillon de cette seconde étape est finalement constitué de 405 patients. Des taux similaires ont été retrouvés dans l'ensemble des pays européens.

Cet échantillon décrit 7,1 % des patients de l'échantillon initial ayant un score bas au GHQ, 25,0 % de ceux ayant un score moyen et 58,6 % des patients ayant un score élevé. 84,4 % des patients ont été revus 3 mois après l'entretien initial et un second suivi à 12 mois est en cours.

LES QUESTIONS POSÉES PAR CETTE ÉTUDE

Nous nous proposons d'envisager une à une les stratégies d'analyse les plus pertinentes pour répondre aux principales questions que soulève une enquête de ce type lorsque doivent être interprétés ses résultats. Il est évident que la construction d'un tel plan d'analyse ne doit pas intervenir *a posteriori* alors que l'enquête est terminée mais doit être prévue initialement dans le protocole expérimental de l'étude. On doit

également souligner que les analyses que nous allons envisager interviennent dans un deuxième temps, alors que les données « brutes » concernant les fréquences des plaintes, les prévalences des diagnostics, les effectifs des différents groupes sociodémographiques, les degrés de handicap, etc., ont déjà fait l'objet d'analyses purement descriptives. C'est sur ce deuxième temps d'analyse que se fonde habituellement la discussion d'un certain nombre d'hypothèses à visées explicatives.

Après avoir recueilli les données brutes de l'enquête et avoir procédé aux premiers tris à plat, aux calculs des effectifs et pourcentages des différents groupes diagnostiques, tout chercheur sera amené à se poser (au moins) les neuf questions suivantes :

1) Les scores obtenus au GHQ sont-ils différents selon les diagnostics?

2) Le nombre de symptômes est-il différent selon les diagnostics?

3) Quel est le degré de corrélation entre le nombre de symptômes (ou les scores au GHQ, ou le score à l'HAMD) et les diagnostics?

4) Quel est l'effet du sexe, de l'âge, du type d'adaptation sociale, d'autres variables familiales ou sociales sur le diagnostic?

5) Existe-t-il une corrélation entre certains diagnostics?

6) Existe-t-il une corrélation entre différents scores et les diagnostics?

7) Existe-t-il une hiérarchie entre les diagnostics?

8) Le type d'organisation (ou la hiérarchie des symptômes) diffère-t-il selon les diagnostics?

9) Sur quel échantillon de population faut-il établir les comparaisons?

Avant d'examiner les solutions les plus appropriées à ces types de problèmes, il convient de faire remarquer d'emblée que les questions 1 à 5 ressortent de l'analyse univariée, les questions 6 à 8 de l'analyse multivariée. Les questions ont été posées dans l'ordre le plus logique sur le plan clinique. Les réponses seront fournies dans l'ordre statistiquement le plus logique.

Questions 1 et 2

Comment tester la différence des scores moyens entre les groupes diagnostiques?

L'approche recommandée lors de la *comparaison multiple de moyennes* est l'*analyse de la variance*, en abrégé : *ANOVA*. En effet, les tests multiples entre différentes paires de moyennes modifient la valeur du coefficient alpha. L'ANOVA protège du risque d'erreur par inflation et permet de savoir s'il existe des différences entre les moyennes des groupes. Si la réponse est *oui*, alors les groupes peuvent être comparés 2 à 2. Dans une ANOVA, lorsque les groupes sont comparés, c'est toujours l'hypothèse nulle qui est testée.

Rappelons le principe d'une analyse de variance à un facteur (en anglais : *one-way* ANOVA). Le terme facteur fait référence à la variable qui constitue les groupes, c'est-à-dire à la variable indépendante (par exemple le sexe si les groupes sont séparés sur ce facteur, ou bien le score à une échelle si un point de coupure a été choisi pour séparer les groupes). Le modèle de l'analyse de variance à un facteur suppose que toute observation individuelle peut être décomposée en :

1) la moyenne générale : $\overline{X}$,

2) le groupe auquel appartient le sujet : *j*,

3) l'observation elle-même : *n*.

Ainsi, si Xnj est la nième observation dans le groupe j et si $\overline{X}j$ est la moyenne du groupe j, $\overline{X}$ étant la moyenne générale des observations, on définit par :

$S(Xnj - \overline{X}j)^2 = SS_E$ la somme du carré des erreurs, et par :

$S(\overline{X}j - \overline{X})^2 = SS_A$ la somme du carré des différences ;

la variance correspond à : SST = SSE + SSA

La formulation la plus souvent employée fait référence à l'interaction du facteur indépendant avec la différence des moyennes retrouvées. Si l'on veut savoir, — comme dans la question que nous nous sommes posée — s'il y a interaction entre le facteur diagnostic (variable indépendante) et les scores moyens, on peut « intuitivement » définir la force de l'interaction F de la façon suivante :

– considérons la variation intragroupe (encore appelée erreur à la moyenne, SEM) qui est la variation entre la valeur affectée à chaque sujet et la moyenne du groupe du sujet ;

– et l'erreur intergroupe ou MSA qui est la variation entre la moyenne de chaque groupe et la moyenne générale X ;

– l'interaction correspond alors à : $F = MS_A/MS_E$.

Il existe des conditions de sensibilité du test F qui sont souvent oubliées lors du recours systématique à l'ANOVA : la distribution des valeurs de la variable dépendante (celle dont les scores sont comparés) doit être par définition normale dans chaque groupe, c'est-à-dire pour chaque « niveau » du facteur. Dans l'exemple choisi, la distribution du score global du GHQ doit être normale chez les hommes et chez les femmes, ou bien chez les patients déprimés et non déprimés (ce qui est facilement obtenu, on le sait, lorsqu'il s'agit de grands échantillons) ; la variance de la population doit être la même dans chaque groupe (on ne compare pas la moyenne d'un score précis sur des populations hétérogènes) ; les observations doivent être indépendantes et l'échantillonnage randomisé.

L'ANOVA à deux facteurs (en anglais *two-way* ANOVA) repose sur le même principe que l'ANOVA à un facteur mais elle prend en compte deux variables indépendantes à la fois. Dans notre exemple, il pourrait s'agir à la fois du statut diagnostique et du statut en terme de score au GHQ (score élevé ou score bas). Plusieurs « niveaux » de facteur sont alors testés (diagnostic A × score bas, diagnostic A × score élevé, diagnostic B × score bas, etc.) et pour chacun la force de l'interaction peut être donnée. Mais l'ANOVA ne peut évidemment fournir aucun renseignement sur le type d'interaction. Lors de l'analyse répétée de la variance, un certain nombre de corrections (la plus connue est celle de Bonferroni) peuvent être apportées pour augmenter la valeur critique de F pour la sensibilité).

Réponse : analyse de variance à un et deux facteurs.

Question 4

Par quelle technique établir l'effet de telle ou telle variable (âge, sexe, etc.) sur l'appartenance à un groupe diagnostique ?

La façon la plus simple d'établir l'effet d'une variable confondante sur la mise en évidence d'une différence entre deux groupes est de procéder à l'« *ajustement* » *des*

groupes sur cette variable, c'est-à-dire de «contrôler» les effets de la variable. Si la différence persiste malgré l'ajustement, alors la différence n'était pas imputable à la variable. Cette méthode est souvent utilisée lorsque des échantillons n'ont pas pu être réellement randomisés et que le rôle d'une variable confondante introduisant un biais doit être éliminé.

Dans notre exemple, nous pouvons ajuster les groupes pour l'âge ou le sexe et vérifier si les différences dans les scores de GHQ persistent (la variable quantitative doit être alors codée en classes) ou si les différences de diagnostic persistent (la variable nominale ici est déjà codée).

La méthode statistique usuelle pour procéder à cet ajustement est l'analyse de la covariance ou ANCOVA. Son principe est d'examiner les variations des valeurs de Y si X est maintenu constant. X est dénommé la covariante puisque l'ajustement est effectué pour les valeurs de X.

L'ANCOVA consiste à calculer la régression de Y par rapport au covariante et à la variable dichotomique codée.

Un autre exemple pour la même étude serait d'ajuster sur le nombre de plaintes (score au GHQ s'il a été retrouvé différent) la différence du nombre de symptômes dépressifs entre déprimés et non déprimés ; ou encore d'ajuster sur le score de GHQ la différence du nombre de symptômes dépressifs entre hommes et femmes (ce qui est certainement une meilleure méthode que la pondération car l'indépendance des observations y est conservée).

Réponse : analyse de la covariance (ANCOVA).

Question 3

Quel est le degré de corrélation entre le nombre de symptômes (ou le score au GHQ) et les diagnostics ?

Le piège d'une telle question réside dans la confusion à ne pas faire entre corrélation de deux variables numériques continues et corrélation entre variable quantitative et classe.

Dans ce dernier cas (qui est celui envisagé par la question), le principe général d'analyse repose sur une régression logistique. La régression logistique est le modèle utilisé quand les variables indépendantes sont indifféremment nominales ou continues et quand la variable dépendante est une variable nominale dichotomique ou catégorielle. Dans l'exemple de référence, la variable nominale dichotomique est le diagnostic et les variables numériques sont soit le nombre de symptômes, soit le score au GHQ, soit un score d'adaptation sociale. En général, la régression se fait pas à pas et un test de chi2 est utilisé pour tester la significativité de l'entrée d'une nouvelle variable.

L'analyse discriminante est possible :

– si la distribution des valeurs des variables indépendantes est une distribution multivariée normale (probablement vrai pour les scores de GHQ et de handicap social) ;

– si la variable dépendante prend au moins deux valeurs.

Ces conditions sont bien remplies dans l'exemple de référence où la variable dépendante correspond à plusieurs diagnostics possibles et où les variables entrant dans la régression sont successivement :

– le nombre de symptômes ;
– les scores du GHQ ;
– les scores de l'HAMD, etc.

Réponse : régression logistique, analyse pas à pas.

Question 6

Existe-t-il une corrélation entre différents scores et les diagnostics ?

Les analyses multivariées doivent être utilisées quand on traite plusieurs variables dépendantes.

Que les variables indépendantes soient numériques ou nominales, une **analyse multivariée de la variance (MANOVA)** est appropriée. Dans notre exemple, les corrélations avec les différents scores (variables dépendantes) seront traitées par diagnostic, chaque score pouvant correspondre à une valeur limite.

Dans le cas particulier où les variables indépendantes sont numériques, on parle plus volontiers d'analyse canonique des corrélations. Mais MANOVA et analyse canonique sont comparatives et apportent peu d'information sur l'organisation des symptômes. La MANOVA est plus souvent utilisée pour comparer l'évolution de scores sous plusieurs traitements. Lorsque l'on désire comparer non pas seulement un nombre de symptômes, mais une organisation symptomatique, le recours aux matrices de corrélation est beaucoup plus informatif.

Réponse : analyse multivariée de la variance (MANOVA).

Questions 7 et 8

Existe-t-il une hiérarchie entre les diagnostics ?

Le type d'organisation (ou la hiérarchie) des symptômes diffère-t-il selon les diagnostics ?

La question n° 7 est parfaitement ambiguë. Si l'on entend par hiérarchie le fait qu'un diagnostic prenne le pas sur les autres lorsqu'il est présent (c'est-à-dire que les autres diagnostics lui sont « secondaires »), tester une telle hypothèse devient très difficile dans la mesure où il s'agit d'une hypothèse de construction. La hiérarchie diagnostique a été prévue par les algorithmes, elle a été prévue lors de la construction même des instruments. Poser la question de la hiérarchie n'a de sens que si l'on se pose la question de la validité de la construction.

Par contre, l'organisation symptomatique elle-même (question n° 8) peut être analysée quant à sa structure latente. Toutes les méthodes d'analyse de structure sont des formes particulières de régression logistique, qu'il s'agisse des *analyses en composantes principales*, des *analyses en cluster*, ou des *analyses du trait latent* (*modèles de Rasch*). Il n'est pas dans notre propos ici de faire une revue critique de ces méthodes. Nous soulignerons seulement que toutes sont des formes d'analyses multivariées reposant sur le principe de l'établissement d'une matrice de corrélation (en l'occurrence la covariation des items pour l'analyse en composantes principales, la covariation des sujets pour l'analyse en cluster, et la covariation des sujets et des items dans l'analyse du trait latent). Le deuxième temps de chacun de ces types d'analyse examine le degré d'information supplémentaire fourni par l'apport de chacune des variables et dérive donc bien des principes d'une régression logistique.

Réponse : analyse du trait latent.

Question 5

Existe-t-il une corrélation entre certains diagnostics ?

Cette question pose le problème de l'association (improprement appelée corrélation) entre variables nominales dichotomiques (ici les diagnostics). La réponse à cette question est étroitement liée à la réponse à la question n° 9 puisqu'il convient d'abord de déterminer quel échantillon de sujets est choisi pour établir l'association des diagnostics : les 2096 patients qui ont passé un GHQ lors de la première étape, ou bien les 405 sujets qui ont répondu au CIDI?

Question 9

Sur quel échantillon de population faut-il établir les comparaisons ?

Dans l'étude de référence, la prévalence des diagnostics a été «projetée» à l'aide de coefficients de pondération sur les 2096 sujets de première étape (c'est-à-dire sur l'ensemble de la population éligible ayant consulté un médecin généraliste). Or, ce qui est licite pour un calcul de prévalence ne l'est plus pour tester statistiquement une association entre valeurs «réelles», c'est-à-dire entre nombre effectif de diagnostics de tel ou tel type. C'est bien sur l'échantillon «réel» de 405 sujets que doit être testée l'association des diagnostics.

La méthode statistique la plus habituelle pour tester l'association entre deux variables dichotomiques consiste à établir un *tableau de contingence* et à calculer le coefficient de contingence du carré de la moyenne. Ce coefficient, appelé *phi de Cramer*, correspond à $\frac{1}{\sqrt{S_1^z + S_2^z}}$. Par construction, il varie entre 0 et 1 puisqu'il ne peut pas avoir de valeurs négatives. Il peut être transformé pour varier entre –1 et +1 et pouvoir ainsi étudier non seulement l'association mais aussi le sens de l'association.

D'autres possibilités qui dépassent le cadre de ce chapitre reposent sur calcul de corrélations tétrachoriques.

Le problème de l'association entre plusieurs diagnostics est souvent posé en termes de «facteur de risque» : soit, si tel diagnostic est présent chez un sujet, quel est le risque supplémentaire que tel ou tel autre diagnostic soit également présent. Nous conclurons donc en examinant brièvement les principes d'analyse des *facteurs de risque* et «*Odds Ratio*».

L'établissement d'un risque relatif n'est possible que dans le cas d'études avec contrôles de cas et suivi clinique. Le plus convaincant serait bien sûr de pouvoir comparer des incidences (nouveaux cas par année) sur de longues périodes. Par exemple, sur deux populations de fumeurs et non fumeurs suivies sur n années, quelle est l'incidence de cancers du poumon? Si on ne peut pas calculer l'incidence, on établit un risque relatif estimé ou «*odds ratio*». Exemple : risque qu'un cas (un cancer) soit fumeur / risque qu'un non-cas (absence de cancer) soit fumeur.

Un autre exemple peut être donné par l'étude de référence : calcul du risque qu'un cas (déprimé) soit une femme / risque qu'un non-cas (non-déprimé) soit une femme.

Ou encore : risque qu'un anxieux soit déprimé / risque qu'un non-anxieux soit déprimé. Comme lors du calcul d'association l'*odds ratio* ne peut être établi que sur des échantillons réels et non pas estimés.

BIBLIOGRAPHIE

DAWSON SAUNDERS B. — *Basic and Theoretical Biostatistics.* Appleton and Lange, 1990.

LAPLANCHE A., COM NOUGUÉ C., FLAMANT R. — *Méthodes statistiques appliquées à la recherche clinique.* Flammarion, Paris, 1991.

LECRUBIER Y., BOYER P., LÉPINE J. P., WEILLER E. — *World Health Organisation Study of Psychological Problems in General Health Care Settings.* Center report : Paris. WHO, Genève, 1993.

LECRUBIER Y., BOYER P., WEILLER E., LÉPINE J. P. — Results of the Paris Center. In : *Mental Illness in General Health Care,* SARTORIUS N., ÜSTUN B. éds, Wiley and Sons, London, 211-225, 1995.

MURPHY E. A. — *Biostatistics in Medicine.* John Hopkins, University Press, Cambridge, 1982.

POCOCK S. J. — *Clinical Trials, a Practical Approach.* Wiley, Chichester, 1983.

SARTORIUS N. *et al.* — An International Study of Psychological Problems in Primary Care. *Archives of General Psychiatry, 50,* 819-824, 1993.

SIMON G., BOYER P., USTUN B., LINDEN M., MAIER W. — Symptom Based and Syndrom Based Diagnostic. *Journal of Affective Disorders* (in press, 1995).

SCHWARTZ D., FLAMANT R., LELLOUCH J. — *L'essai thérapeutique chez l'homme.* Flammarion, Paris, 1980.

SPSS for Windows. User's manual (sur le calcul des facteurs de risque).

15

CLASSIFICATION DES SUJETS ET CLASSIFICATION DES VARIABLES

par S. M. CONSOLI

INTRODUCTION

La *classification automatique*, encore appelée *analyse typologique*, vise à décomposer une population donnée en un certain nombre de sous-groupes homogènes. Les éléments de la population peuvent être des individus, ou des objets, définis par un certain nombre de caractéristiques (valeurs prises par un certain nombre de variables descriptives). C'est essentiellement pour classer des individus ou des objets, que l'analyse typologique est utilisée. Cependant, il peut être également intéressant de l'appliquer à la classification d'un ensemble de variables.

Par convention, on a l'habitude de disposer les individus ou objets dans les lignes d'une matrice (ou tableau) de données, et les variables dans les colonnes de la même matrice; le contenu de la case située au croisement de la i^e ligne et de la j^e colonne représente la valeur prise par la j^e variable chez le i^e individu ou objet. La classification peut donc concerner les lignes, comme les colonnes.

La classification automatique, largement redevable à la puissance de calcul des ordinateurs, n'est que la formalisation mathématique d'une activité cognitive primordiale chez l'être humain, l'activité de *catégorisation*. Nos connaissances sont largement structurées de manière à pouvoir distinguer le *semblable* du *dissemblable* et tout objet de connaissance peut être identifié par la multiplicité de ses appartenances à telle ou telle catégorie.

Synonymes français : *taxinomie, taxonomie, nosologie.*

Synonymes anglais : *cluster analysis, classification, grouping, clumping, Q analysis, numerical taxonomy.*

PROPRIÉTÉS COMMUNES

Il s'agit de *simplifier* une réalité complexe.

Les classes qui seront obtenues sont *inconnues d'avance*. Cette propriété distingue la *classification* d'un ensemble d'objets du *classement* de ces mêmes objets parmi un certain nombre de classes prédéfinies. Choisir un diagnostic parmi une série de diagnostics préétablis est affaire de classement : une fois le diagnostic effectué, le patient ainsi classé sera censé posséder un certain nombre de caractéristiques non

apparentes ou évolutives, en fonction de sa classe d'appartenance (étiopathogénie, pronostic, etc.). Le classement entraîne des connaissances nouvelles sur l'objet qui a été classé mais pas sur l'univers auquel appartient l'objet en question. Reformuler un nouveau découpage des troubles anxieux ou des troubles affectifs, à partir d'un certain nombre de considérations génétiques, biochimiques, sémiologiques, pharmacologiques, épidémiologiques, est affaire de classification. Cette dernière s'accompagne de nouvelles connaissances sur la structuration de l'univers auquel appartiennent les objets à classer. Toutefois, un tel progrès épistémologique dépend de la pertinence des variables choisies pour décrire les objets de connaissance.

Chaque objet ou individu à classer est caractérisé par une *multiplicité de variables ou d'attributs*. De telles variables ou attributs ne sont pertinents pour établir une classification que s'ils apportent une information inégalement répartie d'un objet à l'autre. Par exemple : si tous les objets sont de sexe masculin, la variable sexe n'apporte aucune information utile...

Toutes les variables sont traitées sur un pied d'égalité. Telle variable n'est pas privilégiée aux dépens de telle autre. Il ne s'agit pas d'expliquer une variable ou un groupe de variables («dépendantes») au moyen d'une autre variable ou d'un autre groupe de variables («indépendantes»). L'analyse typologique est une méthode *descriptive* et non *explicative* ou *causale*. Un minimum d'hypothèses est exigé de l'utilisateur.

En revanche, l'*interprétation des classes* obtenues peut déboucher sur de nouvelles hypothèses qui devront être testées à travers de nouvelles observations ou des expérimentations appropriées.

L'analyse typologique poursuit *deux objectifs contradictoires* : obtenir des groupes en *nombre restreint* et *homogènes*. Or, plus le nombre de groupes ou de classes est restreint, plus ces groupes risquent d'être hétérogènes ; toute méthode de classification équivaut à une solution de compromis.

EN QUOI L'ANALYSE TYPOLOGIQUE DIFFÈRE DES AUTRES MÉTHODES D'ANALYSE DES DONNÉES?

L'analyse discriminante

La variable classificatoire et le nombre de classes sont *déterminés à l'avance* (par exemple, la variable de type «*nominal*» ou «*catégoriel*», «personnalité pathologique», en dix classes, ou la variable nominale «pronostic», en trois classes, favorable, incertain, défavorable). On cherche la meilleure combinaison des autres variables descriptives permettant de prédire l'appartenance d'un individu à l'une des classes préétablies : cela suppose que l'on connaisse la valeur de la variable classificatoire («personnalité pathologique» ou «pronostic») pour les individus examinés. À partir de là, tout nouvel individu présentant un certain nombre de caractéristiques pourra être classé dans une de ces catégories avec un risque minimal de classement erroné. L'analyse discriminante et l'analyse typologique peuvent être couplées. On peut chercher quelles sont les variables qui jouent le plus grand rôle dans la constitution des groupes.

L'analyse de segmentation

La variable classificatoire, de type quantitatif (par exemple le score à une échelle de dépression ou à une échelle d'impulsivité) est *fixée par l'utilisateur.* À partir de l'ensemble des autres variables, il s'agit de constituer des groupes d'individus les plus homogènes possibles, par rapport à la variable choisie (par exemple : les déprimés majeurs, les déprimés limites et les non-déprimés; toutefois, les seuils caractérisant chaque segment ne sont pas connus à l'avance).

L'analyse factorielle en composantes principales

Il s'agit de remplacer la multiplicité des variables descriptives (variables quantitatives ou ordinales) par *un petit nombre de facteurs ou dimensions*, indépendants les uns des autres. L'analyse factorielle permet une représentation synthétique des variables. Chaque sujet peut être également caractérisé par ses « scores factoriels », c'est-à-dire par un petit nombre de mesures sur les principales dimensions qui auront été dégagées. La projection des variables sur les plans factoriels aide à l'interprétation des composantes principales. Quant à la projection des individus sur les plans factoriels, elle se visualise généralement sous la forme de « nuages de points » qui peuvent prendre l'allure d'amas plus ou moins denses et homogènes. La lecture des résultats graphiques d'une analyse factorielle invite à procéder à une classification « à vue de nez ». En réalité, les deux méthodes (classification et analyse factorielle) peuvent être associées : les individus projetés sur les plans factoriels peuvent être identifiés par un symbole correspondant à leur classe d'appartenance. Un tel « diagramme de dispersion » peut être complété en traçant sur les plans factoriels des cercles dont les centres coïncident avec les centres de gravité des classes et dont les rayons sont proportionnels à l'écart-type de chaque classe. Cette représentation peut faciliter l'interprétation des résultats d'une analyse typologique.

L'analyse des correspondances

Son principe est analogue à celui de l'analyse factorielle mais les *variables* traitées sont de *type nominal.* Sur les plans factoriels peuvent être projetés les groupes des sujets appartenant à chacune des catégories des variables nominales « actives » (celles qui ont servi à extraire les axes factoriels) mais aussi les catégories de variables « passives » ou « supplémentaires » dont le positionnement peut aider à l'interprétation des résultats (ces variables n'interviennent pas dans l'extraction des axes factoriels). Analyse des correspondances et classification peuvent être couplées : dans ce cas, chaque classe, considérée comme une variable supplémentaire, pourra être projetée sur les plans factoriels, ce qui facilitera son interprétation.

L'analyse des similarités (« multidimensional scaling ») et de préférences (MDPREF)

Ce type d'analyse est destiné à transformer des *mesures non métriques* (classement par ordre de préférence de 10 chansons, ou de 10 présidentiables, par 100 individus) en des distances, qui peuvent être visualisées sur des cartes. Chaque individu peut également être visualisé sur une carte. Ces techniques peuvent être couplées à une analyse typologique. Cela permet alors de regrouper visuellement les individus dont les préférences sont similaires.

À QUOI SERT UNE ANALYSE TYPOLOGIQUE?

Réduction des données

Il s'agit de *tester* les effets de la mise en place d'appartements thérapeutiques dans plusieurs secteurs psychiatriques. On peut, dans un premier temps, essayer de regrouper les 500 secteurs français en trois grandes catégories, à partir d'un certain nombre de caractéristiques (nombre de psychiatres libéraux, nombre de généralistes, nombre de lits en clinique privée, consommation de psychotropes, consommation d'alcool, revenu moyen, taux de chômage, etc.). Les secteurs appartenant à la même catégorie sont considérés comme «similaires». Il faut alors étudier les effets de la mise en place d'appartements thérapeutiques dans trois secteurs psychiatriques, chacun des trois étant sélectionné comme le plus «*représentatif*» de sa catégorie.

Exploration des données

L'utilisateur veut voir plus clair dans une *grande masse de données.*

«– N'ai-je pas oublié une variable descriptive essentielle? (par exemple le sexe des patients, dans la classification de la séméiologie dépressive). Si l'introduction d'une nouvelle variable bouleverse la classification obtenue, c'est que cette variable est sans doute fondamentale et qu'elle avait été oubliée.

– Mon échantillon comporte-t-il des individus marginaux qu'il vaut mieux mettre de côté dans les analyses ultérieures?

– Y a-t-il des sous-groupes relativement homogènes dans mon ensemble volumineux de variables à analyser? Dans ce cas, il serait plus commode d'attaquer l'analyse des données (par exemple par analyses factorielles) en me limitant à des sous-groupes de variables.»

Classification des données

«Je veux *reconnaître* et *nommer des classes* permettant d'avoir une vision plus synthétique que la réalité observée.»

Validation d'hypothèses

«Je veux savoir si mon système de classement (classes définies *a priori*, par exemple en psychopathologie) peut être retrouvé grâce à l'analyse typologique (confirmation ou infirmation). En cas d'infirmation et si je ne pense pas avoir négligé de variables descriptives fondamentales, cela peut me pousser à *réviser mon système de classement.*»

Prédiction fondée sur la nature des groupes

L'analyse typologique peut dégager des groupes de patients dont l'évolution spontanée ou la réponse à une intervention thérapeutique sont différentes. Tout nouvel individu, classé, peut faire l'objet d'une prédiction. C'est ce que l'on fait, intuitivement, lorsque l'on établit un *pronostic.*

Génération d'hypothèses

« Si je suis surpris par la classification obtenue, il va falloir que j'essaie de l'expliquer. Cela me poussera à formuler des hypothèses à vérifier. La structure de la classification (par exemple de la séméiologie dépressive) peut être liée à une variable descriptive que j'avais négligée dans mon observation (par exemple le sexe).

Ou, à l'inverse, j'avais attribué trop d'importance à une variable descriptive qui s'est révélée non informative pour la classification des données.

Ou encore, une "composante" combinant plusieurs variables élémentaires semble rendre compte de la structure des données (par exemple l'anhédonie). Je peux m'efforcer de construire un instrument pour mesurer cette composante. »

QUELLES SONT LES ÉTAPES DE L'ANALYSE TYPOLOGIQUE?

1) Collecte des données, choix des objets et des caractéristiques descriptives ou « variables ». Constitution d'une *matrice de données.*

2) Calcul des proximités (indices de similarité ou de dissimilarité) entre toutes les paires d'objets. Constitution d'une *matrice des proximités.*

3) Utilisation d'un algorithme pour affecter les objets à des groupes (ou classes). Constitution d'une *matrice d'incidence* (d'appartenance au même groupe).

4) Description des groupes, interprétation des résultats. Constitution d'une *matrice des profils de groupes* (chaque groupe est caractérisé par la distribution des variables, par exemple, moyennes et écarts-types, s'il s'agit de variables quantitatives). Tests de liaison (chi-deux, ou analyse de la variance) entre les variables descriptives et les groupes issus de la classification, pour comprendre dans quelle mesure chaque variable descriptive contribue à la constitution des groupes.

5) *Validation des résultats.* Il existe des tests d'« absence de structure » comparant les résultats obtenus avec ceux que l'on obtiendrait si les objets provenaient effectivement d'un groupe unique et homogène. On peut également mesurer le degré de distorsion entre la matrice des similarités initiales et la matrice décrivant l'appartenance des objets aux groupes finaux. On peut enfin tester la validité convergente de deux classifications, obtenues à l'aide de deux algorithmes différents, ou encore de la classification automatique par rapport à un classement basé sur des critères préétablis.

LA COLLECTE DES DONNÉES

Ce stade de la recherche pose en particulier le problème du *choix des variables.* Rappelons qu'il existe quatre types d'échelles de mesure : *nominales* (ou catégorielles), *ordinales* (ou semi-quantitatives), *d'intervalle* et enfin *proportionnelles* (ou métriques). Les deux premiers types correspondent à des variables *qualitatives,* les deux derniers à des variables *quantitatives.* Exemples respectifs : catégorie socioprofessionnelle; réponse à une question sous forme «*pas du tout*», «*un peu*», «*moyennement*», «*beaucoup*» ; température; poids.

Choisir des variables descriptives utiles à la classification implique le fait de *sélectionner des variables informatives,* de retenir des variables du même type

(toutes quantitatives par exemple), faute de quoi il faudra, avant de calculer les proximités entre objets, homogénéiser les variables. On peut aussi (mais cela implique une grande prudence) *pondérer les variables*, c'est-à-dire leur affecter un coefficient, pour privilégier certaines variables ou au contraire pour minorer la part prise par des variables connues avec moins de précision ou mesurant un même concept.

L'*homogénéisation* des variables peut se faire dans le sens descendant (« *appauvrissement* »), par exemple en transformant une variable quantitative en variable ordinale ou une variable ordinale en variable nominale. On peut aussi transformer une variable nominale à plus de deux classes en une série de variables nominales binaires (yeux : bleus, verts, marrons; en yeux bleus : oui/non, yeux verts : oui/non, yeux marrons : oui/non). Dans ce dernier cas, il n'y a pas de perte d'information mais cela peut aboutir à des matrices énormes, difficiles à manipuler.

L'homogénéisation peut aussi se faire dans le sens ascendant (« *enrichissement* »), par exemple en transformant une variable nominale en une variable ordinale. Cela implique une injection d'informations ou l'adjonction d'hypothèses supplémentaires...

Un dernier problème non négligeable, lorsque l'on a affaire à des variables métriques multiples, est celui des *unités de mesure différentes* (par exemple : la taille des individus, exprimée en mètres, et le poids exprimé en kilos). Le calcul des proximités va attribuer une importance excessive aux variables présentant la plus forte dispersion ou variance (dans notre cas, le poids exprimé en kilos). Pour éviter cet inconvénient, l'on conseille de *standardiser* les données, c'est-à-dire d'exprimer ces dernières en nombres d'écarts-types par rapport à la moyenne (toutes les variables ont ainsi une moyenne nulle et une variance égale à l'unité).

CALCUL DES INDICES DE PROXIMITÉ

Un indice de proximité entre deux objets est un nombre unique qui résume les ressemblances ou dissemblances existant sur toutes les variables qui caractérisent ces deux objets.

Il existe deux catégories d'indices de proximité :

– les indices de similarité, s_{ij}, qui mesurent la ressemblance entre les objets i et j;

– et les indices de dissimilarité, d_{ij}, qui mesurent la dissemblance entre les objets i et j.

Ces indices sont :

– des nombres *positifs* ($s_{ij} \geq 0$ et $d_{ij} \geq 0$);

– *symétriques* ($s_{ij} = s_{ji}$ et $d_{ij} = d_{ji}$).

Les indices de similarité sont d'autant *plus élevés* que les paires d'objets se ressemblent davantage ($s_{ij} > s_{ik}$ implique que l'objet i ressemble davantage à l'objet j qu'à l'objet k).

Les indices de dissimilarité sont d'autant *plus faibles* que les paires d'objets se ressemblent davantage ($d_{ij} > d_{ik}$ implique que l'objet i ressemble davantage à l'objet k qu'à l'objet j).

Lorsque l'objet est comparé à lui-même :

$d_{ij} = 0$,

et selon les conventions

$s_{ii} = 1$ ou bien $s_{ii} = s_{\max}$.

Les indices de dissimilarité (mais non les indices de similarité) peuvent vérifier l'*inégalité triangulaire* : $d_{ij} \leq d_{ik} + d_{kj}$, et s'appellent dans ce cas des *distances* (ou distances métriques). Sinon, ils constituent de simples « indices de distance ».

L'espace associé à un indice de dissimilarité qui vérifie l'inégalité triangulaire s'appelle un *espace métrique*. Lorsque l'inégalité triangulaire n'est pas vérifiée (aussi bien pour les indices de similarité que de dissimilarité), on parle d'espace semi-métrique.

Attention ! Toute distance est un indice de dissimilarité, mais tout indice de dissimilarité n'est pas une distance.

Enfin, une propriété particulière sert à procéder à la classification des objets : c'est l'*inégalité ultramétrique* : $d_{ij} \leq \max (d_{ik}, d_{jk})$.

Trois objets dans un espace ultramétrique forment soit un *triangle équilatéral*, soit un *triangle isocèle*. L'objectif de l'analyse typologique hiérarchique est de transformer une matrice de proximités quelconque en une matrice ultramétrique.

En d'autres termes, soit un triangle quelconque ABC avec des côtés de longueur différente AB < BC < CA. Les sommets du triangle représentent trois objets à regrouper. Il va de soi que les deux objets les plus ressemblants (et les moins distants), qui vont être assemblés en premiers, sont A et B. À partir de ce moment, la classe (A, B) sera traitée comme un objet unique : la distance « ultramétrique » entre C et A sera la même que celle entre C et B.

<table>
<tr><td rowspan="4">Indices de proximité</td><td colspan="2">Indices de similarité</td></tr>
<tr><td rowspan="3">Indices de dissimilarité</td><td>Indice de distance</td></tr>
<tr><td>Distance métrique</td></tr>
<tr><td>Distance ultramétrique</td></tr>
</table>

On peut transformer un indice de similarité en indice de dissimilarité et *vice versa* :

Si $s_{ii} = 1$, $d_{ij} = 1 - s_{ij}$ et $s_{ij} = \frac{1}{1 + d_{ij}}$.

Si $s_{ii} = s_{max}$, $d_{ij} = s_{max} - s_{ij}$ et $s_{ij} = \frac{s_{max}}{s_{max} + d_{ij}}$.

On peut aussi transformer un indice de distance (d_{ij}) en distance métrique (d'_{ij})

$d'_{ij} = d_{ij} + C$ avec $C = \max (d_{ij}, d_{jk}, d_{ik})$

Le choix d'un indice de proximité dépend d'abord du type de matrices construites lors de la collecte des données.

Matrices de mesures

Chaque *ligne* = un objet ; chaque *colonne* = une variable quantitative d'intervalle ou de proportion.

• ***Distance euclidienne*** : racine carrée de la somme des carrés des différences (théorème de Pythagore) ; on peut aussi travailler sur des distance euclidiennes au carré.

Inconvénient : distance biaisée en direction des variables qui ont la plus grande dispersion. On peut pallier cet inconvénient *en standardisant les variables* ou en pondérant chaque variable par l'inverse de sa variance (on obtient ainsi le carré d'une distance euclidienne). Autre inconvénient de la distance euclidienne : distance biaisée en direction des objets qui ont la plus grande « taille » (c'est-à-dire les valeurs les plus élevées sur l'ensemble des variables).

Imaginons deux crânes adultes, un longiligne et l'autre bréviligne et deux crânes d'enfants, un longiligne et l'autre bréviligne. Le recours à la distance euclidienne, même après standardisation des variables (largeur et hauteur du crâne), risque d'assembler les deux crânes d'adulte et les deux crânes d'enfant et de passer à côté de l'opposition longiligne/bréviligne. On peut pallier cet inconvénient par une double pondération : on pondère les objets par l'inverse de leur « *taille* » (somme des valeurs de toutes les variables pour un objet) et les variables par l'inverse de leur « *importance* » (somme des valeurs de cette variable pour tous les objets).

• Cette distance s'appelle la ***distance du chi-deux***. Elle ne peut être calculée que sur une matrice de données positives.

Inconvénients de ces distances : elles assument implicitement que les variables sont indépendantes. Si une dimension est mesurée par plusieurs variables redondantes (fortement corrélées entre elles) et si une autre dimension n'est mesurée que par une seule variable, il y aura surestimation de la dimension surreprésentée. Certains auteurs conseillent de procéder dans un premier temps à une analyse factorielle en composantes principales et de travailler dans un deuxième temps sur la matrice des scores factoriels.

• Une autre méthode, équivalente, est de recourir à la ***distance de Mahalanobis***, calculée sur les données brutes. Cette distance pondère chaque paire de variables par l'inverse de leur covariance.

Inconvénient de la distance de Mahalanobis : elle réduit les distances entre objets appartenant à des groupes différents et augmente les distances entre objets du même groupe, ce qui est exactement l'inverse de l'objectif de l'analyse typologique !

Toutes ces distances sont calculées à partir des carrés des différences de mesure, ce qui conduit à accorder une importance maximale aux objets atypiques.

• Si l'on désire privilégier les petites différences, on peut recourir à la ***distance rectangulaire*** ou « *city block* » ou *distance de Manhattan* (la valeur de l'hypoténuse d'un triangle rectangle est égale à la somme des côtés à angle droit).

• *On peut enfin calculer directement un indice de similarité sur une matrice de mesures.* Il suffit de calculer le **coefficient de corrélation r_{ij} de Pearson** entre les objets i et j. Le principe est exactement le même que pour le calcul de la corrélation entre deux variables quantitatives. Dans l'exemple donné plus haut, le calcul d'un coefficient de corrélation entre objets permettrait de classer ensemble les crânes longilignes (de l'adulte et de l'enfant) et, dans un deuxième groupe, les crânes bréviliges.

Attention ! Ce coefficient varie entre –1 (dissemblance maximale) et +1 (ressemblance maximale). Ce n'est donc pas à proprement parler un indice de proximité (qui doit toujours être positif). On peut, à partir du coefficient de corrélation, passer à un

indice de distance, par la formule $d_{ij} = 1 - r_{ij}$ ($d_{ij} = 0$ si corrélation parfaite ; $d_{ij} = 1$ si corrélation nulle ; $d_{ij} = 2$ si corrélation négative parfaite). Cet indice est intéressant pour assembler des variables. Il est critiquable pour assembler des objets car il dépend de l'échelle de mesure des variables.

Matrices d'occurrence ou de fréquence

Pour une matrice d'occurrence : chaque *ligne* = un objet ; chaque *colonne* = une caractéristique ; contenu d'une case : nombre de fois où la caractéristique *j* a été attribuée à l'objet *i* (par exemple, on demande à 100 sujets si 10 adjectifs qualificatifs décrivent bien 20 présidentiables : la matrice comportera 20 lignes et 10 colonnes et chaque case contiendra un nombre compris entre 0 et 100).

Une matrice de fréquence est une matrice d'occurrence dont le contenu de chaque case a été divisé par le total de la matrice.

• La distance la plus adaptée à ce cas est la ***distance du chi-deux***.

• On peut aussi utiliser un indice de similarité fondé sur la théorie de l'information, appelé ***information mutuelle***.

Matrices de rang

Chaque *ligne* = un objet ; chaque *colonne* = une variable ordinale.

Par exemple, on demande à 50 psychiatres de classer par ordre de « tolérance » ou d'« efficacité » dix antidépresseurs. La matrice comporte 50 lignes et 10 colonnes. Par convention, en cas d'*ex aequo*, on calcule un rang moyen : par exemple si le 4e et le 7e andidépresseurs ont été classés *ex aequo* en deuxième position, on leur attribue le rang 2,5.

• À partir de là, on peut calculer une distance euclidienne entre objets,

• ou bien, une distance rectangulaire,

• ou bien un coefficient de corrélation de rang de Spearman (coefficient rho) qui variera entre –1 et +1,

• ou bien le coefficient de similarité tau de Kendall, variant de 0 à 1.

Matrices logiques

C'est ce qu'on appelle également un *tableau de codage disjonctif*, ou un *tableau de présence-absence*.

Une *ligne* = un objet ; une *colonne* = une des modalités d'une variable nominale.

Les variables nominales sont également appelées « attributs ».

Chaque case contient un 0 et un 1.

Deux objets « partagent » un certain nombre d'attributs (co-présences P), « partagent » un certain nombre d'absences d'attributs (co-absences A), et « divergent » au niveau des autres attributs (non-coïncidences N) sur un total de T attributs.

Diverses formules existent selon l'importance ou la pondération que l'on accorde aux co-présences, co-absences et non-coïncidences.

COMMENT CONSTITUER DES GROUPES À PARTIR D'UNE MATRICE DES PROXIMITÉS?

L'objectif de toute classification est de simplifier l'information contenue dans une matrice des proximités en remplaçant les proximités entre paires d'objets par des proximités entre groupes d'objets.

Les critères de classification dépendent de l'objectif du chercheur (quelle structure de classification souhaite-t-il obtenir?), de la nature de l'information traitée (matrice des données/matrice des proximités) et de l'étendue de l'information manipulée (critère «local» avec traitement de l'information de proche en proche, ou critère «global»).

Plusieurs grands types de classifications peuvent être obtenus :

– *Une partition* : ensemble de parties de l'ensemble initial d'objets, mutuellement disjointes et dont la réunion est égale à l'ensemble initial d'objets : tout objet appartient à une classe et à une seule.

– *Une hiérarchie de partitions* (ou arbre de partitions); on part de la partition la plus fine (un objet = une classe) et on assemble progressivement des classes jusqu'à obtenir l'ensemble d'objets tout entier. Une fois classé dans une branche de l'arbre, un objet ne peut pas changer de branche, à un niveau de partition plus grossière.

– *Un recouvrement* : ensemble de parties non disjointes de l'ensemble initial d'objets et dont la réunion est égale à l'ensemble initial d'objets : un objet peut appartenir à deux classes (ou plus de deux classes) à la fois.

– *Une hiérarchie de recouvrements* : même principe que pour la hiérarchie de partitions, en sachant qu'un objet peut appartenir à deux classes à la fois.

– *Une famille de partitions* : pas de structure hiérarchique : ensemble de partitions p_1, p_2,...p_i, indicées par le nombre de classes (1 seule classe pour p_1, 2 classes pour p_2, etc.).

– *Une famille de recouvrements* : pas de structure hiérarchique non plus.

Méthodes monothétiques

La grande majorité des méthodes de classification est de type *polythétique* (on tient compte de *toutes les variables à la fois*). Il existe cependant aussi des méthodes *monothétiques* qui ne tiennent compte que d'*une seule variable à la fois* : à chaque étape, la division se fait selon une variable différente; ces méthodes sont hiérarchiques, divisives et séquentielles. Elles s'appliquent à des matrices logiques (ou matrices d'attributs).

Deux critères classiques :

– *l'analyse des associations de Lambert et Williams*. Elle consiste à calculer les chi-deux entre un attribut et tous les attributs restants et à faire la *somme de tous ces chi-deux*. On refait la même chose pour chacun des attributs. Le critère de classification est celui qui maximise une telle somme. Cela revient à choisir, pour la première division de l'ensemble des objets, la *variable* (ce peut être une question d'un questionnaire avec des réponses types oui/non) *la plus sélective*;

– un critère proche, mais plus exact, consiste à calculer l'*information mutuelle entre paires d'attributs* (entropie de l'attribut *i* expliquée par l'attribut *j*).

Méthodes polythétiques

Elles se subdivisent en *méthodes hiérarchiques*, qui permettent d'obtenir une famille de partitions, et en méthodes non hiérarchiques.

Les méthodes hiérarchiques se subdivisent elles-mêmes en méthodes *ordinales*, qui n'utilisent comme information que le classement des paires d'objets par ordre de proximité, et en méthodes *non ordinales*, qui utilisent les valeurs numériques des proximités entre paires d'objets. La plupart des méthodes utilisées recourent à un *critère local* (la proximité entre deux groupes n'est fonction que des proximités entre les objets constituant ces deux groupes) ; il existe cependant aussi des méthodes qui recourent à un critère global (la constitution des groupes fait intervenir la totalité des proximités).

Parmi les méthodes les plus connues, on peut citer :

– critère de plus proche voisin, et critère du voisin le plus éloigné, pour des méthodes ordinales ;

– méthode de la moyenne, méthode centroïde, méthode de la médiane, et méthode de Ward, pour les méthodes non ordinales.

En ce qui concerne les méthodes non hiérarchiques, elles sont nombreuses mais comprennent deux grandes catégories :

– *Méthodes de réallocation*, où un objet affecté à un groupe donné lors de la i^e itération peut être réaffecté à un autre groupe lors d'une itération ultérieure de l'algorithme. Exemples : méthode des K-means, et méthodes des nuées dynamiques. Ces méthodes aboutissent à une partition des objets.

– *Méthodes de recherche de densité*, par exemple : l'analyse modale de Wishart. Ces méthodes aboutissent à une classification des objets en « groupes denses », mais c'est à l'utilisateur de définir, en observant les résultats, le nombre de groupes qu'il décide de choisir, en sachant que des « objets isolés » peuvent ne pas se trouver affectés aux groupes denses.

Assemblage simple ou méthode du plus proche voisin, appelée également chaînage simple (single linkage or nearest neighbor, or minimum method)

Les deux premiers objets (lignes ou colonnes, selon ce que l'on veut agréger) qui sont agrégés, sont ceux dont la distance est la plus petite, parmi l'ensemble de toutes les distances entre objets pris deux à deux (matrice de distances). Ce couple est alors considéré comme un nouvel objet. Les distances entre ce nouvel objet (à deux unités) et tous les autres objets sont calculées et définies comme les distances minimales entre les objets individuels et chaque unité de la classe à deux objets. Et ainsi de suite. On peut considérer les enveloppes ou périphéries des nuages constitués par les nouveaux assemblages (ce sont des classes en cours de constitution). La distance d'un nouvel objet à un nuage sera la plus petite distance à l'enveloppe de ce nuage.

Inconvénient de cette méthode : elle donne des résultats peu satisfaisants lorsque des objets intermédiaires sont présents entre deux groupes ou lorsque les groupes ne sont pas nettement séparés (effet de « chaînage »).

Assemblage complet ou méthode du voisin le plus éloigné, appelée encore chaînage complet (complete linkage or furthest neighbor or maximum method)

La distance entre deux nuages est calculée comme la distance entre leurs points les plus éloignés. Cette méthode produit des nuages « compacts » et de taille (nombre d'objets par classe) comparable.

Méthode de la moyenne (average linkage between groups method)

La distance entre deux nuages est la moyenne des distances de toutes les combinaisons possibles de paires d'objets appartenant à la fois à l'un et à l'autre nuage. Une variante de cette méthode, appelée «*average linkage within groups*», assemble deux nuages en s'arrangeant pour que la moyenne des distances entre toutes les paires possibles au sein du nuage résultant, soit la plus petite possible. La méthode «par la moyenne» est souvent la méthode, utilisée «par défaut», par les logiciels.

Méthode centroïde ou du centre de gravité

La distance entre deux classes est définie comme la distance entre leurs centres de gravité.

L'inconvénient de cette méthode est que, à force d'agrégations, les distances entre nuages finissent par décroître, ce qui peut paraître paradoxal.

Autre inconvénient : si une classe est de taille très faible, l'information qu'elle apporte risque d'être perdue.

Méthode de la médiane

Proche de la méthode centroïde. Ici, pour le calcul du nouveau centre de gravité, on donne le même poids aux deux nuages qui vont être agrégés, quels que soient les nombres d'objets qu'ils contiennent, respectivement. Cela évite de sous-estimer les classes de petite taille.

Méthode de Ward (ou critère du moment centré d'ordre 2)

Le moment centré d'ordre 2 d'une classe d'objets est égal à la somme du carré des écarts des objets au centre de gravité de la classe. Il est aussi égal à la somme des carrés des distances euclidiennes entre toutes les paires d'objets de la classe, divisées par le nombre d'objets. Cela revient à calculer une distance pondérée entre centres de gravité. L'intérêt de ce critère par rapport à la méthode centroïde est que les distances entre classes vont en croissant, au fur et à mesure que les classes sont fusionnées.

Attention : les méthodes de la «moyenne», «centroïde», «médiane», et de «Ward», ne peuvent s'employer qu'avec un carré de la distance euclidienne. Les méthodes d'assemblage «simple» ou «complet» peuvent aussi bien procéder à partir de matrices de similarité, type corrélation r de Pearson.

Il faut savoir aussi que toutes les méthodes, en dehors de l'assemblage «simple», de l'assemblage «complet», et de la méthode de «Ward» peuvent donner, en classification hiérarchique ascendante, des branches qui ne se connectent jamais aux autres : cela est dû au fait qu'à partir d'un certain stade, la distance ou dissimilarité requise pour agréger les deux classes suivantes décroît au lieu de croître à chaque étape.

Méthode des K-means ou méthode de Mac Queen

Cette méthode permet d'obtenir une *partition unique des objets en k classes*, le nombre k ayant été spécifié *a priori* par l'utilisateur ou pouvant être déterminé par la méthode elle-même.

On choisit k premier objets de l'échantillon pour constituer les k premiers *centres de groupes*, mais il est aussi possible de tirer au hasard les k objets. Une première étape de l'algorithme génère une configuration initiale en k classes autour des k noyaux. Une deuxième étape procède à une réallocation éventuelle de chaque objet à une autre classe que sa classe initiale d'appartenance, de façon à obtenir k groupes *les plus homogènes possible* (l'algorithme procède par itération jusqu'à ce que les centres de gravité soient stables).

L'intérêt de cette méthode est *d'épargner le temps de traitement informatique* lorsque l'on a un nombre important d'objets à classer (plus de 100). Les données doivent être des distances euclidiennes.

Méthode des nuées dynamiques de Diday

Elle est proche de la méthode de Mac Queen.

Méthodes de recherche de densité

Ces méthodes sont basées sur le calcul d'une densité affectée à chaque objet à classer. La densité est définie comme la distance moyenne des k objets les plus rapprochés de l'objet considéré. k est fixé par l'utilisateur : habituellement, il est compris entre 3 et 5. Lorsque k est égal à 1, la méthode de recherche de densité est équivalente à la méthode du voisin le plus proche.

L'algorithme consiste à classer dans un premier temps l'ensemble des n objets par ordre de densité décroissante. On considère ensuite qu'il existe un seul groupe, identifié à l'objet de densité maximale. Les autres objets sont soit affectés à un groupe dense déjà existant, soit constitués eux-mêmes en nouveaux groupes denses. On obtient ainsi progressivement un ensemble de nuages denses, avec des points isolés (non denses) extérieurs à ces nuages. L'étape finale correspond au regroupement de tous les objets.

On peut visualiser un tel procédé dans un espace à deux dimensions en traçant des courbes d'équidensité, tout à fait comparables aux courbes d'altitude d'une carte géographique destinée à représenter le relief. Les groupes denses sont équivalents à des groupes montagneux séparés par des cols ou des vallées, la courbe d'altitude la plus basse regroupant la totalité de l'espace géographique concerné.

Cette méthode travaille sur des distances euclidiennes. Elle a l'intérêt de ne pas être trop sensible à la forme des nuages (elle peut parfaitement identifier des nuages qui ne seraient pas convexes) et de pouvoir distinguer entre objets typiques, appartenant à des classes denses, et objets marginaux, dont on peut ne pas tenir compte dans une classification.

INTERPRÉTATION ET VALIDATION DES RÉSULTATS

Il est d'usage de *représenter graphiquement* les classifications hiérarchiques (familles de partitions sous forme d'*arbres* allant de la partition la plus fine (un

objet = une classe) à la partition la plus grossière (classe unique englobant la totalité des objets)). Les classes qui s'agrègent sont comme les rameaux et les branches d'un arbre, les niveaux de fusion ou d'agrégation correspondant à la valeur de l'indice de distance ou de l'indice de dissimilarité à partir de laquelle l'algorithme a créé une nouvelle classe.

On peut également *dresser « les profils » des classes*, en utilisant les variables actives (celles qui ont servi à la classification) ou des variables passives, qui n'intervenaient pas dans la description initiale des données. Ces profils peuvent être représentés *graphiquement*, si l'on a affaire à des variables quantitatives (sur le mode bien connu des résultats du MMPI par exemple). Des *tests appropriés* peuvent permettre de repérer les variables discriminantes (chi-deux, test de Student, analyse de la variance, voire analyse discriminante).

On peut enfin coupler l'analyse typologique et une autre forme d'analyse des données comme une *analyse factorielle en composantes principales* ou une *analyse factorielle des correspondances* et dessiner les nuages de points correspondant à la partition choisie en k classes sur le plan factoriel formé par les deux premiers axes.

Nous n'avons fait que mentionner les méthodes de validation des résultats. Divers auteurs ont mis au point des tests appropriés, notamment pour évaluer la distorsion que l'on fait subir aux données en réduisant l'information initiale lorsque l'on classe les objets. D'autres formes de validation existent, ne serait-ce que la vérification de la stabilité de la structure obtenue avec divers critères de classification ou encore de la *convergence* entre le résultat d'une classification automatique par rapport à un critère préétabli de classement.

En réalité, la meilleure validation d'une classification automatique est la mise à l'épreuve expérimentale des hypothèses qu'elle permet de générer et notamment des prédictions sur le devenir des objets en fonction de leur classe d'appartenance (sensibilité à un traitement, pronostic, découverte d'une étiologie différente d'une classe à l'autre).

Bibliographie

Ouvrages et articles généraux

BARTHELEMY J. P., GUENOCHE A. — *Les arbres et les représentations des proximités*. Masson, Paris, 1988.

BENZECRI J. P. — *L'analyse des données. Tome I : la taxonomie*. Dunod, Paris, 1973.

BERTIER P., BOUROCHE J. M. — *Analyse des données dimensionnelles*, 2e édition. PUF, Paris, 1977.

BOUROCHE J. M., SAPORTA G. — *L'analyse des données*. PUF (Coll. Que sais-je?), Paris, 1980.

CHANDON J. L., PINSON S. — *Analyse typologique. Théories et applications*. Masson, Paris, 1981.

DUCIMETIÈRE P. — Les méthodes de classification numérique. *Revue de statistique appliquée*, *18*, 4, 5-26, 1970.

JAMBU M., LEBEAUX M. O. — *Classification automatique pour l'analyse des données. Tome I : Méthodes et algorithmes; Tome II : Logiciels*. Dunod, Paris, 1978.

LERMAN I. C. — *Les bases de la classification automatique*. Gauthiers Villars, Paris, 1970.

ROUX M. — *Algorithmes de classification*. Masson, Paris, 1985.

Quelques applications à la psychiatrie

FALS-STEWART W., LUCENTE S. — An MCMI Cluster Typology of Obsessive-Compulsives : a Measure of Personality Characteristics and its Relationship to Treatment Participation, Compliance and Outcome in Behavior Therapy. *J. Psychiatr. Res.*, *27*, 139-54, 1993.

HAM L. P., ANDRASIK F., PACKARD R. C., BUNDRICK C. M. — Psychopathology in Individuals with Post-Traumatic Headaches and other Pain Types. *Cephalalgia*, *14*, 118-26, 1994.

HANNAH M. T. — An Empirical Typology of Seriously and Persistently Mentally Ill Patients using Symptom and Social Functioning Factors. *J. Clin. Psychol.*, *49*, 622-37, 1993.

HORNER T. N., Jr., UTERMOHLEN V. — A Multivariate Analysis of Psychological Factors Related to Body Mass Index and Eating Preoccupation in Female College Students. *J. Am. Coll. Nutr.*, *12*, 459-65, 1993.

MOWBRAY C. T., BYBEE D., COHEN E. — Describing the Homeless Mentally Ill : Cluster Analysis Results. *Am. Community Pyschol.*, *21*, 67-93, 1993.

RILEY J. L., 3d, ROBINSON M. E., GEISSER M. E., WITTMER V. T. — Multivariate Cluster Analysis of the MMPI-2 in Chronic Low-Back Pain Patients. *Clin. J. Pain.*, *9*, 248-52, 1993.

STEER R. A., BECK A. T., BROWN G. K., BECK J. S. — Classification of Suicidal and Nonsuicidal Outpatients : a Cluster-Analytic Approach. *J. Clin. Psychol.*, *49*, 603-14, 1993.

SZATMARI P., BOYLE M. H., OFFORD D. R. — Familial Aggregation of Emotional and Behavioral Problems of Childhood in the General Population. *Am. J. Psychiatry*, *150*, 1398-403, 1993.

TALMINEN T. J., HELENIUS H. — Suicide Clustering in a Psychiatric Hospital with a History of a Suicide Epidemic : a Quantitative Study. *Am. J. Psychiatry*, *151*, 1087-8, 1994.

YOSHINO A., KATO M., TAKEUCHI M., ONO Y., KITAMURA T. — Examination of the Tridimensional Personality Hypothesis of Alcoholism using Empirically Multivariate Typology. *Alcohol Clin. Exp. Res.*, *18*, 1121-4, 1994.

16

L'APPROCHE DE LA PERSONNALITÉ PAR LES TRAITS

par M. HUTEAU

La psychologie de la personnalité poursuit un double objectif : décrire et expliquer des noyaux de cohérence entre les conduites, noyaux de cohérence qui sont relativement stables et qui témoignent d'une différenciation interindividuelle (Huteau, 1985). La cohérence entre conduites est d'autant plus intéressante qu'elle porte sur des conduites nombreuses, variées et distantes. Elle peut prendre des formes diverses : manifestation de conduites identiques dans des situations voisines, corrélations fortes entre la présence de certaines conduites dans certaines situations et celles de conduites toutes différentes dans des situations toutes différentes également, évolutions de nature différente en fonction de la structuration initiale des conduites. Les noyaux de cohérence, qui peuvent être désignés comme des invariants personnels, peuvent être appréhendés de diverses manières : à partir des buts poursuivis, des motifs de l'activité, des opérations intellectuelles privilégiées, des règles d'interprétation du monde, des schémas de soi... Formellement, la personnalité est le plus fréquemment décrite en termes de traits ou de types.

TRAITS ET TYPES

Les traits sont des variables latentes au long desquelles il est possible d'ordonner les individus. Selon le niveau auquel il se situe dans le trait, le sujet tend à manifester des comportements particuliers ou des comportements généraux d'une certaine intensité. Cette position du sujet sur le trait est généralement considérée comme une disposition interne relativement stable (Buss, 1989). Au niveau descriptif, on distingue des traits bipolaires (les individus se répartissent entre deux pôles) et des traits unipolaires (les individus sont plus ou moins proches d'un pôle). Plutôt que de chercher à mettre en évidence d'emblée un trait bipolaire, il est souvent préférable de considérer deux traits unipolaires (chacun correspondant à un pôle du trait bipolaire envisagé) et de s'interroger sur leur corrélation (le trait bipolaire est alors fondé si les deux traits unipolaires sont en forte corrélation négative). À un autre niveau, plus explicatif, on distingue des traits de source et des traits de surface. Les traits de surface sont de simples résumés descriptifs d'un ensemble de caractères qui, dans la vie courante, covarient. Les traits de source, moins nombreux, sont mis en évidence dans des situations épurées et permettent d'expliquer les traits de surface.

Les types sont des catégories disjointes censées regrouper des individus qui se ressemblent. La catégorie peut être définie par le sujet le plus proche de la moyenne des individus qui la composent ou par le sujet le plus typique de cette catégorie (on parle alors de type idéal). Très fréquemment, et c'est une tradition qui remonte à la caractérologie du XIX[e] siècle, les types sont construits à partir du croisement des traits : après avoir établi une partition sur les traits (en général une dichotomie), on caractérise les individus par leur profil sur l'ensemble des traits considérés. On peut citer comme exemple la typologie de Heymans et Wiersma élaborée au début du siècle et popularisée en France par Lesenne et Berger. Les sujets sont ventilés en 8 types qui résultent du croisement des traits dichotomisés : émotivité, activité et retentissement (primarité-secondarité). Une manière plus moderne, mais relativement peu utilisée, de définir des types à partir des traits consiste à représenter les sujets par des points dans un espace multidimensionnel dont les dimensions sont des traits, et à considérer que les types existent si les points représentant les sujets constituent des amas homogènes. On considère généralement que la description de la personnalité au moyen de types est trop schématique dans la mesure où elle ne tient pas compte du caractère peu typique de nombreux individus. Il n'en demeure pas moins que certains questionnaires de personnalité fournissant des descriptions sous forme de types rencontrent un succès appréciable. Tel est le cas du *Myers-Brigg Type Inventory* (MBTI) fondé sur la typologie de Jung qui distingue 16 types à partir du croisement des dichotomies, introverti ou extraverti, préférence pour le jugement ou la perception, l'intuition ou la sensation, la pensée ou le sentiment (Huteau, 1989).

LA MISE EN ÉVIDENCE ET L'INTERPRÉTATION DES TRAITS

Comment savoir si un trait dont on considère l'existence, qu'il s'agisse ou non d'une catégorie du sens commun, a une réalité effective? Pour répondre à cette question, on doit d'abord expliciter la définition du trait hypothétique. Il faut ensuite faire correspondre à cette définition un ensemble de situations dans lesquelles le sujet ne doit pas se conduire de la même manière selon qu'il occupe une position ou une autre sur le trait. Ces situations peuvent être des situations réelles; on observe alors la fréquence ou l'intensité des comportements réels du sujet. Elles peuvent aussi être des situations évoquées; les conduites des sujets dans ces situations sont alors décrites par des personnes les connaissant bien ou par les sujets eux-mêmes (comme dans les questionnaires de personnalité). On considérera que le trait existe si la cohérence attendue entre les conduites dans les diverses situations retenues se manifeste bien. Cette cohérence est généralement estimée par des coefficients de corrélation qui indiquent la probabilité de se comporter d'une certaine manière dans une situation lorsqu'on a manifesté tel comportement dans une autre. Il y a isomorphisme entre le modèle des traits et celui de l'analyse factorielle : si dans toutes les situations prises deux à deux les conduites covarient, c'est que toutes ces situations sont saturées dans un même facteur. Ce facteur est un trait. Le trait étant défini par l'ensemble des situations où il se manifeste, sa signification s'enrichira au fur et à mesure que l'on montrera qu'il contribue à l'explication de la conduite dans de nouvelles situations. À partir d'hypothèses théoriques sur la nature du trait, on pourra construire des situations spécialement destinées à éprouver ces hypothèses. Si celles-ci sont validées, la signification du trait sera encore enrichie.

Illustrons l'ensemble de cette démarche de mise en évidence et d'interprétation des traits à partir de travaux classiques de H. J. Eysenck sur le trait bipolaire introversion-extraversion qui est, avec l'anxiété, le trait largement le plus étudié (Eysenck, 1967, 1970, 1990; Huteau, 1983). Pour Eysenck, les sujets plutôt extravertis se caractérisent notamment par leur sociabilité, leur impulsivité, leur hyperactivité et leur excitabilité. Les sujets plutôt introvertis ont des caractéristiques inverses. À partir de cette définition, on peut construire toute une série d'items de questionnaires qui présentent des situations où l'on peut manifester ou non sociabilité, impulsivité, activité ou excitabilité. L'analyse statistique des réponses des sujets montre que plus les sujets sont sociables, plus ils ont tendance à être impulsifs, actifs et excitables dans des situations diverses. On peut alors considérer que le trait existe et qu'il est légitime d'ordonner les sujets sur un continuum introversion-extraversion (défini par exemple à partir du nombre de réponses données témoignant de conduites extraverties). Tous les constats réalisés au moyen du questionnaire ainsi construit pourront contribuer à renforcer et à préciser la signification du trait. On a observé par exemple que les sujets extravertis bénéficiaient davantage des méthodes d'éducation actives au cours des formations (les sujets introvertis bénéficient plus des méthodes traditionnelles) et que dans la vie professionnelle leur absentéisme était plus fréquent. On a observé aussi que les troubles hystériques étaient plus fréquents chez les extravertis et les troubles obsessionnels plus fréquents chez les introvertis, ou encore, et plus généralement, que la désadaptation prenait plus fréquemment la forme de la délinquance chez les extravertis et celle de la névrose chez les introvertis.

Eysenck a présenté une théorie biologique de l'introversion-extraversion dans laquelle il propose d'expliquer les différences de comportement des introvertis et des extravertis par la plus grande sensibilité et la plus grande activité spontanée du cortex cérébral des premiers. L'activation du cortex cérébral est largement déterminée par le système réticulaire ascendant et, pour Eysenck, c'est le degré d'activité de ce système réticulaire, plus exactement de la boucle réticulo-corticale, qui détermine le degré d'introversion. À partir de cette proposition, on peut prédire des différences de comportements dans toute une série de situations expérimentales selon que les individus sont plutôt introvertis ou extravertis. C'est ainsi que l'on peut s'attendre à ce que les sujets introvertis aient une réaction électro-dermale plus forte, ou encore soient davantage capables de maintenir leur attention. D'assez nombreux résultats sont compatibles avec la théorie d'Eysenck. On peut noter que si la description du trait peut être assez proche des définitions communes, son explication nous en éloigne. L'homme de la rue peut très bien décrire l'introversion-extraversion en termes de sociabilité et d'impulsivité, mais il est peu probable qu'il nous explique pourquoi les sujets introvertis se conditionnent plus rapidement que les extravertis lorsque les stimuli sont faibles et que c'est l'inverse lorsqu'ils sont forts!

LES GRANDES DIMENSIONS DE LA PERSONNALITÉ

On peut s'intéresser à des traits particuliers choisis pour leur intérêt pratique ou théorique. On peut aussi, et c'est plus ambitieux, s'intéresser à l'ensemble des traits de personnalité et à son organisation. Pour définir l'ensemble des traits, dans cette perspective de description exhaustive, il serait nécessaire de disposer de la liste de l'ensemble des comportements possibles. On pourrait alors, à partir des covariations observées entre comportements, mettre en évidence les traits constituant la

personnalité. On ne voit pas très bien comment il serait possible d'établir la liste de tous les comportements, liste qui, de toute manière, serait provisoire. Pour tourner cette difficulté, on considère, depuis les années 1930 avec Allport et Odbert, que les termes de la langue permettant de se décrire et de décrire autrui constituent le répertoire complet des comportements. On peut alors utiliser ces termes pour décrire le comportement des sujets, cette description pouvant être faite par les sujets eux-mêmes ou par des observateurs les connaissant. Bien que discutable, car elle suppose que la connaissance psychologique est contenue dans la langue et que l'on peut y accéder par condensations et extractions successives, cette démarche s'est révélée heuristique.

Les recherches de R. B. Cattell, conduites dans cette perspective dès les années 1940, ont mis en évidence 12 traits de base ou facteurs primaires lorsque les sujets sont décrits par autrui et 16 traits lorsqu'ils se décrivent eux-mêmes au moyen de questionnaires. Ces traits n'étant pas indépendants, Cattell a montré qu'on pouvait ramener cette liste à un nombre plus restreint de grands traits ou facteurs secondaires (Cattell et Kline, 1977). Dans le prolongement de ce travail pionnier, les nombreux travaux réalisés ont convergé vers un modèle comportant 5 dimensions ou traits bipolaires larges : le modèle des « big five » qui est très proche des propositions de Cattell (John, 1990; Goldberg, 1993). Voici la liste de ces cinq dimensions, les pôles de chacune sont illustrés par trois adjectifs.

1. Introversion-extraversion (E)
   pôle E : bavard, actif, énergique
   pôle I : tranquille, réservé, timide
2. Amabilité positive-amabilité négative (A)
   pôle A + : sympathique, gentil, flatteur
   pôle A – : culpabilisant, froid, inamical
3. Consciencieux/peu consciencieux (C)
   pôle C + : organisé, minutieux, prévoyant
   pôle C – : insouciant, désordonné, frivole
4. Stabilité émotionnelle-anxiété (N)
   pôle N + : tendu, anxieux, nerveux
   pôle N – : stable, calme, satisfait
5. Ouverture aux expériences (O)
   pôle O + : champ d'intérêts large, imaginatif, original
   pôle O – : champ d'intérêts étroit, banal, simple

Il y a bien sûr une part de convention dans le choix de ne retenir que cinq dimensions même si cette structure est retrouvée dans des cultures différentes et avec diverses modalités d'observation. Aussi le large consensus qui s'est réalisé pour juger cette description pertinente est-il assez étonnant et il est vraisemblable qu'il n'est que passager. Ces dimensions constituent aujourd'hui un cadre de référence quasiment obligatoire dans lequel on tend à situer systématiquement toutes les descriptions de la personnalité.

Certes, on peut penser aussi qu'il est excessivement réducteur de limiter à cinq le nombre des points de vue que l'on peut prendre pour examiner la variabilité interindividuelle. En fait, ces cinq dimensions doivent être vues comme l'étage supérieur d'un modèle hiérarchique comportant au moins trois niveaux. Chaque grande

dimension du modèle «coiffe», ou résume, des traits de niveau inférieur, traits que l'on appelle parfois des facettes. C'est ainsi, par exemple, que l'introversion-extraversion comporte, dans le questionnaire de Costa et McCrae (1985) les facettes suivantes : niveau d'activité, dominance, sociabilité, absence de contrôle, émotion positive. On retrouve ainsi, mais plus clairement dissociés, les composants de l'introversion-extraversion signalés par Eysenck. Chacun de ces traits (le niveau d'activité, la dominance,...) peut à son tour coiffer des traits de niveau inférieur. Cette représentation de l'organisation des traits montre bien que la personnalité peut être décrite avec des degrés de précision différents. Au niveau le plus élevé de la hiérarchie, on a la description la plus synthétique mais aussi la plus grossière. Aussi, à des fins d'applications pratiques, en clinique notamment, aura-t-on généralement intérêt à considérer des traits se situant à un niveau subordonné par rapport aux «*big five*» (Petot, 1994).

DISPOSITIONS ET SITUATIONS : LES POINTS DE VUE INTERACTIONNISTES

Les conduites des individus, et c'est en cela qu'elles sont adaptatives, se caractérisent à la fois par la stabilité de leur organisation (leur cohérence) et par leur flexibilité. En même temps qu'elles témoignent de l'existence de dispositions internes, qui constituent la personnalité et qui sont conceptualisées le plus souvent sous forme de traits, les conduites se modifient en fonction des exigences des situations, ce dont rendent bien compte les diverses théories de l'apprentissage. Les difficultés d'articulation de ces deux propriétés fondamentales de l'organisme, ou en d'autres termes la difficulté à penser simultanément la permanence d'un ensemble et la mobilité de ses parties, ont conduit certains à valoriser excessivement le poids des dispositions internes tandis que d'autres valorisaient tout aussi excessivement le poids des facteurs situationnels. Pour une certaine personnologie, le rôle des situations est négligeable : la conduite s'explique par les dispositions internes. Pour un certain situationnisme, c'est l'inverse et la conduite s'explique exclusivement par les circonstances. On peut distinguer deux variantes du situationnisme. Avec le situationnisme radical qui s'inspire du behaviorisme skinnerien, il n'y a pas de différences individuelles stables, donc pas de personnalité au sens habituel du terme. Les différences entre sujets ne sont que le reflet des circonstances (en l'occurrence les contingences de renforcement) et elles peuvent être facilement modifiées en changeant ces circonstances. Un situationnisme moins radical, souvent cognitif dans la mesure où il met l'accent sur la perception de la situation, admet l'existence de différences individuelles stables, mais on considère que celles-ci sont étroitement associées à des situations spécifiques et qu'il n'y a guère de sens à les conceptualiser sous forme de traits (*cf.* par exemple Cantor et Kihlstrom, 1987).

Ces divergences ont suscité à la fois des débats, des polémiques et des travaux expérimentaux (Huteau, 1985; Angleteiner, 1991; Endler et Parker, 1992; Magnusson, 1990; Mischel, 1990). Il ressort de ces travaux que les deux positions précédentes sont difficiles à tenir. La cohérence de la conduite a été souvent surestimée : les covariations entre conduites observées relevant d'un même trait sont souvent faibles, de l'ordre de .30 (ceci n'est pas incompatible avec une stabilité élevée des différences individuelles : lorsqu'on applique le même questionnaire de personnalité aux mêmes sujets à plusieurs années d'intervalle, on observe

fréquemment des corrélations de l'ordre de .60-.70). Cette cohérence est parfois illusoire dans la mesure où elle est en partie construite par l'observateur. En effet, en observant autrui, ou en nous observant nous-mêmes, et si de strictes précautions méthodologiques ne sont pas prises, nous avons fortement tendance à sélectionner les observables et à interpréter les ambiguïtés afin de donner du sujet ou de nous-mêmes une image unitaire. Si la prédiction de la conduite à partir des traits n'est pas très bonne, celle qui se fonde sur les situations n'est pas meilleure. Très fréquemment, les individus placés dans une même situation se comportent différemment, c'est d'ailleurs la raison pour laquelle il existe une psychologie différentielle. La difficulté à opter pour les facteurs situationnels ou pour les facteurs dispositionnels a conduit au développement de points de vue interactionnistes où ces deux classes de facteurs sont envisagées simultanément. On peut distinguer au moins trois points de vue de ce genre.

1) La manière la plus simple d'envisager cette interaction consiste à raisonner en termes d'effet et à se demander si l'effet du trait sur une conduite est ou non le même selon la situation, ou, ce qui revient au même, si l'effet de la situation est ou non le même selon la position occupée par l'individu sur le trait. À l'effet de la disposition et à celui de la situation s'ajoute alors un effet d'interaction généralement massif. (Dans cette perspective, on donne au terme interaction la signification qu'il a en statistique : il y a interaction lorsque l'effet d'une variable varie selon les modalités d'une autre variable.)

2) Le second point de vue conduit à une redéfinition des traits. Si les conduites censées relever d'un même trait ne sont que faiblement associées, c'est vraisemblablement qu'elles concernent un ensemble de situations trop hétérogène. On se propose donc bien de définir des traits, mais relativement à une classe spécifiée de situations. Cette démarche peut être illustrée à partir des travaux sur l'anxiété. Dans beaucoup de théories, celle-ci est conceptualisée comme un trait bipolaire général (c'est le cas, nous l'avons vu, dans le modèle des « *big five* ») et il existe de nombreux questionnaires permettant de caractériser les sujets par leur degré d'anxiété. Les recherches de N. S. Endler *et al.* (1991, par exemple) ont montré qu'il était possible de distinguer quatre classes de situations dans lesquelles l'anxiété est susceptible de se manifester plus ou moins fortement. Les trois premières classes sont fondées sur des travaux factoriels où l'on montre qu'il y a trois facteurs d'anxiété qui correspondent à des situations où l'on est l'objet d'une évaluation sociale, des situations où l'on court un danger physique et des situations inconnues ou ambiguës. La quatrième classe de situations représente les situations habituelles de la vie quotidienne qui ne sont pas spécialement anxiogènes mais dans lesquelles l'anxiété peut néanmoins se manifester. Plutôt que de caractériser les sujets par un trait « anxiété générale », on les caractérise donc par quatre traits. Les corrélations entre ces traits sont faibles (de l'ordre de .30-.40 entre les trois premiers traits, de l'ordre de .00-.20 entre chacun de ces trois premiers traits et le quatrième). La démarche suivie par Endler mériterait certainement d'être généralisée à l'ensemble des traits de personnalité.

3) Le 3ᵉ point de vue interactionniste consiste à proposer des modèles fonctionnels où le processus d'élaboration de la conduite comporte une série de paramètres, certains relatifs aux situations, d'autres relatifs à des dispositions internes. M. Bruchon-Schweitzer (1994),par exemple, rend compte des stratégies d'adaptation à la maladie en évoquant des facteurs situationnels comme les événements stressants dans leurs composantes objective et subjective ou le soutien social dont bénéficie le

malade et des facteurs dispositionnels comme l'anxiété-trait, l'optimisme ou le style d'attribution causale (contrôle interne ou externe). Dans leurs formes les plus sophistiquées, ces modèles fonctionnels font appel à des processus de rétroaction et d'autorégulation : la conduite du sujet a des effets sur les paramètres qui ont contribué à son apparition. Ces modèles, qui pour certains sont les seuls modèles vraiment interactionnistes (on parle aussi de modèles transactionnels), s'ils rendent bien compte de l'influence des situations et des dispositions sur la conduite, permettent aussi de comprendre comment ces situations et dispositions sont susceptibles d'être modifiées (l'individu ne subit pas seulement l'influence des situations, il les transforme et les choisit aussi ; s'il subit bien l'influence de ses dispositions, son activité contribue aussi à les infléchir), en ce sens ce sont des modèles dynamiques.

CONCLUSION

Les traits occupent toujours une place centrale dans la psychologie de la personnalité et on continue à les prendre en compte dans les divers secteurs de la pratique psychologique. Si les traits ont parfois été considérés dans le passé comme des entités abstraites, on pense aujourd'hui qu'il y a tout intérêt à les définir relativement à des classes de situations et à envisager leur action dans le cadre de modèles du fonctionnement psychologique en interaction avec des facteurs associés aux situations dans lesquelles les sujets sont impliqués.

BIBLIOGRAPHIE

ANGLEITNER A. — Personality Psychology : Trends and Development. *European Journal of Personality*, *5*, 185-197, 1991.

BRUCHON-SCHWEITZER M. — Les problèmes d'évaluation de la personnalité aujourd'hui. *L'Orientation Scolaire et Professionnelle*, *23*, 35-57, 1994.

BUSS A. H. — Personality as Traits. *American Psychologist*, *44*, 1378-1388, 1989.

CANTOR N., KIHLSTROM J. F. — *Personality and Social Intelligence*. Prentice-Hall, Englewood Cliffs N. J., 1987.

COSTA P. T. Jr, MC CRAE R. R. — *The Neo Personality Inventory Manual*. Psychological Assessment Ressources, Odessa, FL, 1985.

CATTELL R. B., KLINE P. — *The Scientific Analysis of Personality*. Academic Press, New York, 1977.

ENDLER N. S., PARKER J. D. A., BAGHY R. M., COX B. J. — Multidimensionality of State and Trait Anxiety : Factor Structure of the Endler Multidimensional Anxiety Scales. *Journal of Personality and Social Psychology*, *60*, 919-926, 1991.

ENDLER N. S., PARKER J. D. A. — Interactionism Revisited : Reflexions on the Continuing Crisis in the Personality Area. *European Journal of Personality*, *6*, 177-198, 1992.

EYSENCK H. J. — *The Biological Basis of Personality*. Charles C. Thomas, Springfield, IL, 1967.

EYSENCK H. J. — *The Structure of Human Personality* (3rd Ed). Methuen, Londres, 1970.

EYSENCK H. J. — Biological Dimensions of Personality. In : L. A. Pervin (Ed.), *Handbook of Personality, Theory and Research*, Guildford, New York, 1990.

GOLDBERG L. R. — The Structure of Phenotypic Personality Traits. *American Psychologist*, *48*, 26-34, 1993.

HUTEAU M. — Les questionnaires de personnalité HSPQ et EPI. *L'Orientation Scolaire et Professionnelle*, *12*, 243-260, 1983.

HUTEAU M. — *Les conceptions cognitives de la personnalité*. PUF, Paris, 1985.

HUTEAU M. — Étude de la validité interne du MBTI (forme G). *Revue de Psychologie Appliquée*, *39*, 101-118, 1989.

JOHN O. P. — The «Big Five» Factor Taxonomy : Dimensions of Personality in the Natural Language and in Questionnaire. In : L. A. Pervin (Ed.), *Handbook of Personality, Theory and Research*, Guilford, New York, 1990.

MAGNUSSON D. — Personality Development from an Interactional Perspective. In : L. A. Pervin (Ed.), *Handbook of Personality, Theory and Research*, Guilford, New York, 1990.

MISCHEL W. — Personality Dispositions Revisited and Revised : a View after Three Decades. In : L. A. Pervin (Ed.), *Handbook of Personality, Theory and Research*, Guilford, New York, 1990.

PETOT J. M. — L'intérêt clinique du modèle de personnalité en cinq facteurs. *Revue Européenne de Psychologie Appliquée*, *44*, 57-63, 1994.

17

L'ÉVALUATION DE LA QUALITÉ DE VIE EN PSYCHIATRIE

par Ph. CIALDELLA et O. CHAMBON

L'ÉMERGENCE DU CONCEPT DE QUALITÉ DE VIE

Les interrogations sur la recherche du bonheur et du bien sont très anciennes. Peraut (1992) rappelle le point de vue soutenu par deux célèbres écoles philosophiques : l'école stoïcienne, pour laquelle le bonheur constitue le but suprême mais inaccessible, et qui propose que l'homme échappe à la souffrance de cette frustration en distrayant son esprit de la quête de ce bonheur. Pour l'école épicurienne, en revanche, le seul objectif de l'homme est la recherche du plaisir.

En dehors de ces aspects philosophiques, on peut décrire au concept de Qualité de Vie (QV) une triple origine sociale (et politique), psychologique et médicale (*Health-Related Quality Of Life*, HRQOL).

Il est devenu classique de souligner que c'est le président Johnson qui a lancé le terme de QV, lors d'un discours (1964) : «Les objectifs [de notre politique] ne peuvent pas être évalués seulement en fonction de la balance des paiements. Ils ne peuvent se mesurer qu'à la qualité des vies menées par nos concitoyens.» L'apparition de cette préoccupation est probablement liée aux conditions sociopolitiques de l'époque qui ont favorisé un changement de valeurs dans l'esprit des citoyens des pays développés : accroissement du niveau de revenus et de confort, naissance des préoccupations écologiques et consuméristes et d'une contestation de la société de consommation, allongement de la durée de vie.

Les travaux initiaux sur la QV proviennent de sociologues : on parle du «mouvement des indicateurs sociaux», avec les grandes enquêtes de Bradburn *et al.* (1969), d'Andrews et Withey (1976), et de Campbell (1976), citées par Peraut (1992). Des indicateurs objectifs de QV comme les revenus, l'éducation, la santé, l'emploi, la criminalité, la participation politique, la croissance démographique, le niveau de revenus, etc., ont été étudiés, mais il a été très tôt reconnu que la QV subjective avait plus de valeur.

Campbell *et al.* (1976) ont ainsi fait remarquer qu'entre 1957 et 1972, alors que la plupart des indicateurs économiques et sociaux progressaient rapidement, la proportion des Américains qui se déclaraient «très heureux» décroissait régulièrement. Un grand nombre d'études devait démontrer l'absence de corrélations entre indicateurs objectifs et vécu de bonheur ou de bien-être subjectif. D'autres travaux, cités par Diener (1984), montrèrent que les relations entre indicateurs sociaux et bien-être

subjectif étaient plus complexes, et variaient selon les conditions d'observation, d'expérimentation, et de mesure.

Par la suite, les recherches sociologiques ou psychologiques sur la QV ou le bien-être subjectif (Diener, 1984), ont accordé une grande importance à la notion de satisfaction par rapport aux domaines de vie. Ces travaux sont à rapprocher de la théorie de la motivation selon Maslow (1943), pour lequel les besoins de l'homme doivent être satisfaits de façon hiérarchique : à la base se trouvent les besoins physiologiques, puis ceux de stimulation et de protection, d'appartenance et d'amour, d'estime de soi et des autres, et, enfin, les besoins de réalisation de soi.

Les premières références à la QV dans le champ médical proviennent de la définition de la santé proposée par l'OMS (1947), à une époque de progrès médical rapide. Par la suite, on devait constater un fléchissement notable des progrès de la médecine en termes d'années de vie gagnées, en dépit des augmentations importantes des budgets de la santé. L'allongement de l'espérance de vie a fait apparaître que de nombreuses maladies étaient chroniques, invalidantes et nécessitaient des soins prolongés dont il fallait évaluer l'impact.

En parallèle, apparaissait une médecine préventive, visant à diminuer les facteurs de risque chez des sujets *a priori* non malades (par exemple les personnes souffrant d'hypertension artérielle non compliquée), par des interventions (traitements, régimes, etc.) dont les conséquences sont parfois négatives sur le vécu du patient (contraintes, effets secondaires, etc.).

La problématique du choix entre plusieurs thérapeutiques fournit un autre paradigme de l'application du concept de QV en médecine (ex. : une chirurgie risquée mais curatrice ou un traitement médical moins risqué mais non curateur?), car on n'envisage plus seulement le nombre d'années de vie gagnées, mais aussi la qualité des années de vie gagnées.

Le domaine médical reste marqué par la prépondérance des approches fonctionnelles (performances objectives ou états de santé objectivables) malgré l'apparition de mesures plus subjectives, centrées sur la santé perçue. Le concept de QV subjective a le mérite de redonner la parole au patient, qui avait été un peu oublié par des médecins un peu trop fascinés par la toute-puissance de la technique.

L'attrait pour le concept de QV est devenu général comme en témoigne le flot de publications y faisant référence. L'industrie pharmaceutique se préoccupe de plus en plus de fournir des évaluations de la QV comme critère de jugement dans les essais médicamenteux, et la FDA (*Food and Drug Administration*) a inclus l'évaluation de l'impact sur la QV dans le cahier des charges des évaluations des nouveaux médicaments.

Les psychiatres sont probablement mieux armés que leurs collègues des autres spécialités pour intégrer le concept de QV dans leurs discours, car la psychiatrie est, dans le champ de la médecine, la discipline qui accorde le plus d'importance au discours et aux perceptions du patient. Par ailleurs, les psychiatres étaient déjà familiers de ce domaine depuis les années 60 avec les difficultés de construction des échelles et des auto-questionnaires, ce qui nous donne une certaine avance conceptuelle.

DÉFINITIONS ET MODÈLES THÉORIQUES

Il n'existe pas actuellement de consensus sur une définition claire et précise de la QV, d'autant que le concept se situe au carrefour des sciences économiques et sociales, de la médecine, et de la psychologie.

Cependant, depuis les travaux sur les indicateurs sociaux qui devaient démontrer la relative indépendance entre indicateurs objectifs et vécu de bien-être, la très grande majorité des auteurs s'accorde à reconnaître que la QV est de nature psychologique et subjective, et échappe aux évaluations normatives.

Pour Joyce (1987), l'évaluation de la QV implique un jugement de valeur, et il affirme que « lorsque les valeurs sont envisagées, celles qui m'importent sont les miennes », pointant l'aspect subjectif de cette évaluation et de ces normes.

Une première grande distinction doit être opérée entre les définitions dites « cognitives » de la QV et les définitions « affectives ». Campbell *et al.*, cités par Corten (1992) ont ainsi différencié la notion de satisfaction, synonyme selon eux, de QV, et celle de bien-être (*well-being*) : la satisfaction repose sur un jugement par rapport à des objectifs ou besoins personnels, ou par rapport à autrui (comparaison sociale) et implique une dimension cognitive, alors que le bien-être reflète une dimension affective, émotionnelle, qui est plus de l'ordre du « ressenti ».

D'autres définitions plus récentes de la QV, comme celle de Calman (1984), rappelée par Zannotti (1991), mettent également l'accent sur une approche cognitive-évaluative : « La qualité de vie mesure la différence, à un moment donné, entre les attentes et espoirs d'un individu et son expérience présente. » De même pour la définition de Gerin *et al.* (1991) : « La qualité de vie se définit comme l'ensemble des satisfactions/insatisfactions éprouvées par un sujet à propos de sa vie en général. »

D'autres auteurs font plutôt référence au bien-être, et donc à la dimension affective : ainsi en est-il de la définition de la santé proposée par l'OMS en 1947, avant l'apparition du terme de QV : « La santé n'est pas seulement l'absence de maladie ou d'infirmité mais un état total de bien-être physique, mental et social. » À l'inverse du concept positif de « bien-être », mais en restant dans le champ affectif, certains auteurs ont proposé de définir la QV comme l'absence de détresse émotionnelle (stress, désespoir, insécurité, etc.) ou d'inconfort psychologique.

L'approche en termes de satisfaction (ou de distance par rapport à des objectifs) est probablement plus stable dans le temps que celle en terme de bien-être, qui a tendance à suivre de plus près les mouvements de l'humeur et des émotions de l'individu, ou la réactivité aux événements de vie.

D'une façon générale, la plupart des auteurs s'accordent à reconnaître que la QV est multidimensionnelle ou multifactorielle, mais les points de vue diffèrent quant à la façon de concevoir les dimensions ou facteurs. Pour Andrews et Withey (1976), trois facteurs indépendants composent le bien-être subjectif : le jugement personnel sur la vie en général, les affects positifs et les affects négatifs. En 1980, Andrews et McKennel (1980) ne reconnaîtront plus qu'une seule composante affective bipolaire associée à la composante cognitive-évaluative.

Une deuxième approche de la multidimensionnalité consiste à distinguer des domaines de vie ; généralement pour évaluer la satisfaction de l'individu dans chacun de ces domaines. Certains auteurs tentent de calculer des scores de QV globale à partir de sommes de satisfactions par domaines, tandis que d'autres préfèrent utiliser

des profils de satisfaction, s'interdisant de calculer un indice global de QV (Gerin *et al.*, 1991).

La délimitation des domaines de vie est assez variable selon les auteurs : la typologie la plus simple, inspirée de la définition de la santé de l'OMS, distingue le physique, le psychologique et le social. Joyce (1987) a développé une typologie plus détaillée, mais encore assez simple des domaines de vie connue par le sigle PCASEE pour *Physical, Cognitive, Affective, Social, Economic, Ego-functioning*, et à laquelle Bech (1993), accorde une grande importance dans sa définition de la QV.

D'autres types de délimitation des domaines ont été proposés, notamment dans le champ de la santé. Par exemple, Lehman (1982, 1988, 1990) utilise 8 domaines de vie qui sont le lieu de vie, la famille, les relations sociales, les loisirs, le travail, la sécurité, l'argent, la santé. Cette dernière approche n'est pas directement superposable à l'approche PCASEE. L'exemple du travail est assez illustratif : dans le modèle PCASEE, le domaine « travail » ne se retrouve pas directement, mais peut être abordé du point de vue physique, social, économique, et du point de vue de la réalisation personnelle (*ego-functioning*). Certains auteurs tendent à complexifier les domaines de vie, tandis que d'autres préfèrent les simplifier. Ainsi, pour Gerin *et al.* (1991) qui regroupent les divers domaines de leur questionnaire PQVS en un facteur « fonctionnel » (gênes physiques) et un facteur « relationnel ».

Des critiques ont été avancées à l'encontre de l'approche par domaines : les domaines pointés par les différents auteurs ne sont pas toujours comparables, et ne revêtent pas la même importance selon l'âge, le sexe, la culture, etc. On a avancé que la QV globale résultait d'une intégration reposant sur une pondération par hiérarchisation des domaines de vie, les domaines valorisés par le sujet ayant un impact plus important sur le vécu de bien-être, que ceux qui le sont moins.

Gerin *et al.* (1991) ont proposé de pondérer les jugements de satisfaction par domaines de vie, en fonction des importances accordées à ces domaines. Cependant, ces auteurs utilisent des mesures d'importance absolue par domaine, plutôt que des importances relatives (hiérarchies de valeurs).

En dehors du classique modèle ascendant (*bottom-up*) dans lequel les domaines de vie contribuent au bien-être subjectif ou à la perception de satisfaction globale par le canal des satisfactions partielles, on décrit aussi un modèle descendant (*top-down*), dans lequel c'est le bien-être subjectif global qui contribue à la perception de satisfaction par domaines (Diener, 1984). En faveur du modèle descendant, on sait que la QV globale est très influencée par les troubles de l'humeur et notamment la dépression, comme l'ont montré Abbey et Andrews (1984).

L'évaluation de la QV ne reste pas confinée à la sphère subjective, et les indicateurs objectifs (conditions de vie, ressources, etc.) définissent pour de nombreux auteurs un domaine de QV objective, reconnu comme dépendant des normes sociales et culturelles locales. Lehman (1982) associe dans son modèle multidimensionnel de QV, une composante objective, une composante subjective et des caractéristiques personnelles.

Le terme « objectif » peut avoir une autre signification lorsqu'il est employé pour insister sur le comportement du sujet, notamment en termes d'autonomie et de relations sociales. Le grand problème des approches objectives vient du fait qu'il n'existe pas de norme universelle des comportements.

Qu'en est-il des définitions de la QV dans le domaine de la santé? On sait déjà par les travaux du mouvement des indicateurs sociaux, que l'influence de la santé sur la

QV globale ne semble se manifester que pour des atteintes sérieuses et durables de la santé (Diener, 1984). De nombreuses définitions médicales de la QV se sont intéressées à la mesure de la santé ou des gênes fonctionnelles dues à la maladie, aux effets secondaires et aux contraintes des traitements. Certaines insistent sur les états de santé objectivables, et d'autres sur les états de santé subjectifs, comme la définition de Schipper *et al.* (1990) pour lesquels la QV représente « l'effet fonctionnel d'une maladie et de son traitement sur un patient, tel que perçu par le patient ». Pour l'équipe de Gerin et Dazord (1991), la plupart des échelles dites de QV dans le champ médical ne sont en fait que des échelles de santé, c'est-à-dire de performances, ou à l'inverse, de gêne et de handicap.

Une autre approche intéressante (Weinstein, 1992), mais critiquée, est celle des *Quality Adjusted Life Years* ou QALYs. Du fait des contraintes économiques, il est parfois nécessaire de faire des compromis entre qualité et quantité de vie gagnée. On parle alors d'analyse d'utilité, méthode appartenant au cadre général de l'analyse de décision, qui traite des valeurs et des préférences dans de telles situations. Pour de nombreux patients, dix années de vie d'une bonne qualité sont préférables à quinze années de mauvaise qualité. Le principe revient à définir une pondération des années de vie, par une estimation de la QV (au sens de la santé) sur une échelle allant de 100 % (santé parfaite) à 0 % (mort).

Pour établir les valeurs de pondération correspondant aux différents états de santé intermédiaires entre la santé parfaite et la mort, il faut d'abord définir les états de santé puis évaluer les préférences du patient pour ces états de santé. La méthode QALY prend donc en compte elle aussi les évaluations subjectives pour l'estimation des valeurs d'utilité, mais reste une méthode de mesure d'état de santé, plus que de QV.

Enfin, une question insistante de l'évaluation de la QV dans le domaine de la santé consiste à savoir s'il faut privilégier des instruments généraux non spécifiques d'une pathologie (moins sensibles mais plus proches du concept de QV, et permettant des comparaisons entre pathologies) ou des instruments spécifiques par pathologie (plus sensibles, mais ne reflétant plus que des aspects très partiels de la QV).

Le sens du terme de QV varie donc selon les études et les auteurs. Comme nous avons besoin de définitions, même arbitraires, nous proposons que le terme « QV » soit assimilé à la satisfaction globale dans les domaines de vie (ou bien à l'inverse de la distance entre les objectifs de vie et l'état de l'individu par rapport à ces objectifs). On pourrait réserver le terme de « bien-être » aux aspects affectifs ou émotionnels, et le terme de « performances/gênes » ou « d'ajustement » aux manifestations plus comportementales. QV et bien-être sont forcément subjectifs et nécessitent des auto-évaluations, tandis que performances/gênes et ajustement peuvent être évalués à la fois par le sujet et par des observateurs.

PROBLÈMES GÉNÉRAUX DES INSTRUMENTS D'ÉVALUATION DE LA QV

La plupart des instruments de QV s'intéressent à la subjectivité du patient, mais il ne faut pas confondre approche subjective et auto-évaluation. Une auto-évaluation est souvent utilisée pour connaître les ressources objectives d'un patient (logement,

argent, travail, etc.). Inversement, on peut recueillir des opinions subjectives au travers d'entretiens structurés, les réponses étant codées par un observateur extérieur en fonction de grilles, ce qui constitue une hétéro-évaluation.

Les auto-évaluations sont plus proches du vécu du patient, mais comportent l'inconvénient d'une fiabilité douteuse, en raison de l'impact de la désirabilité sociale, du déni ou du mensonge volontaire (Zautra et Goodhart, 1979). Les hétéro-évaluations obtenues au travers d'entretiens structurés comportent le risque de biais d'interprétation de la part de l'investigateur, et peuvent se heurter à des résistances de la part du sujet.

Au niveau des items, on remarquera que les items unipolaires repris d'échelles de souffrance ou de gêne ne sont souvent pas bien adaptés à l'évaluation de la QV, car l'évaluation de la QV nécessite aussi d'évaluer le bien-être et la satisfaction, qui ne se résument pas à l'absence de souffrance ou de gêne. Aussi, est-il préférable de disposer d'items bipolaires, explorant à la fois le pôle de bien-être ou de satisfaction et celui de détresse ou d'insatisfaction.

Dans tous les questionnaires, une attention particulière doit être portée au libellé des questions, afin d'éviter les incompréhensions, les biais de réponse dus à des formulations complexes (par exemple pour les questions négatives, ou pour les questions explorant simultanément plusieurs phénomènes au sein d'un même item), ou à des questions trop ressemblantes dans leur formulation.

Enfin, comme cela a été démontré avec les items portant sur les affects, il faut parfois prudemment distinguer les questions portant sur la présence, l'intensité, la fréquence, ou la proportion relative de durée des phénomènes.

Certains instruments sont excessivement simples et se contentent d'un seul item, pour une évaluation de la QV ou du bien-être globaux, tandis que d'autres sont composés d'un très grand nombre d'items, parfois organisés en domaines. Les instruments peuvent devenir assez lourds et complexes, comme le questionnaire PQVS (Dazord *et al.*, 1993) qui comporte plusieurs domaines et plusieurs dimensions par domaine (gêne, satisfaction, importance, attente de changement, adaptation à la gêne) ainsi que deux versions différentes pour les essais thérapeutiques : une version pour l'inclusion et une pour l'évaluation du résultat.

Les instruments d'évaluation doivent satisfaire à des conditions métrologiques suffisantes de fiabilité et de validité. La *Food and Drug Administration* américaine (selon Shumaker *et al.*, 1990) exige cinq conditions : linéarité sur toute l'étendue mesurée (homogénéité), sensibilité au changement, signification clinique des changements observés (validité de contenu et de construct), validité de construct ou transférabilité (identité de valeur des scores dans différents troubles, ou à différentes périodes), validité divergente d'avec les échelles mesurant les symptômes cliniques ou les effets secondaires.

Ces conditions sont rarement remplies et l'examen de la validité est rendu très ardu par la complexité des questionnaires de QV et des modèles multidimensionnels sous-jacents.

APPLICATION DE L'ÉVALUATION DE LA QUALITÉ DE VIE EN PSYCHIATRIE

La qualité de vie (QV) des patients souffrant de pathologies psychiatriques suscite, depuis le début des années 80, un intérêt croissant auprès des chercheurs et des cliniciens en santé mentale.

Évaluation de la QV des patients dits « névrotiques »

Par « névrotiques », nous entendons les pathologies avérées où prédominent les manifestations anxieuses et/ou dépressives, sans altération profonde du rapport à la réalité.

La méthodologie d'évaluation de la QV dans ce type de troubles ne semble pas présenter de différences majeures par rapport à celle de patients souffrant de pathologies physiques, dans la mesure où des auto- et/ou hétéro-évaluations peuvent être utilisées facilement, et où le sujet a, en général, libre accès à sa propre subjectivité, même si l'*insight* peut être très limité.

Un petit nombre d'instruments spécifiques aux troubles névrotiques ont été développés récemment :

– l'adaptation du questionnaire PQVS pour les patients dépressifs (Dazord *et al.*, 1993) ;

– le *Quality of Life Enjoyment and Satisfaction Questionnaire* de Endicott *et al.* (1993), qui comporte 93 questions.

Certains laboratoires pharmaceutiques travaillent sur des instruments d'évaluation de la QV pour névrotiques : ainsi l'échelle du laboratoire SmithKlineBeecham, SBQOL (Dunbar *et al.*, 1989). Dès les années 70, plusieurs instruments d'évaluation de l'adaptation sociale pour les patients névrotiques avaient été publiés, dont certains comportaient des sections que nous qualifierons actuellement de QV ; ainsi les échelles d'adaptation sociale de Katz et Lyerly (1963), le *Structured and Scaled Interview to Assess Maladjustment* (SSIAM) de Gurland *et al.* (1972), et l'inventaire de Clare et Cairns (1978) cités par Cialdella (1992). Ces questionnaires ont finalement été assez peu utilisés en psychiatrie.

Quant aux instruments généraux d'évaluation de la QV, applicables théoriquement à toutes les pathologies, ils ne sont pas toujours bien adaptés à l'évaluation des problèmes de santé des névrotiques : l'échelle de Kaplan, le *Sickness Impact Profile* (SIP), et le *Nottingham Health Profile* (NHP) sont plutôt des indicateurs de santé objective ou perçue. Le questionnaire de McMaster et l'*Anamnestic Comparative Self-Assessment* (ACSA) de Bernheim semblent plus correspondre à la définition que nous avions retenue, mais n'ont pas, à notre connaissance, été utilisés pour les troubles névrotiques ou dépressifs, dans leurs versions complètes.

Certains instruments ont été conçus dans le contexte de l'évaluation des services psychiatriques dans la communauté (Bigelow *et al.*, 1982 ; Baker et Intagliata, 1982) et sont donc nettement plus orientés vers les patients psychotiques (*cf.* paragraphe suivant).

Dans certaines publications récentes sur la QV des patients dépressifs (Wells *et al.*, 1991) ou anxieux (Feighner, 1987 ; Markowitz *et al.*, 1989), les auteurs ont plutôt centré leur propos sur la détresse affective subjective, sur les effets secondaires des

traitements, ou sur les conséquences fonctionnelles, professionnelles et sociales des troubles (Weissman, 1981), mais assez peu sur la composante « satisfaction » de la QV.

Par ailleurs, il est intéressant de noter que de nombreux instruments utilisés pour mesurer l'intensité de la dépression, de l'anxiété ou de la souffrance psychologique générale chez les patients névrotiques sont couramment utilisés comme échelles de QV ou de *well-being* en dehors du domaine psychiatrique (Shumaker *et al.*, 1990). Le principal inconvénient de ces échelles vient du fait que les items sont souvent unipolaires et ne mesurent que les phénomènes pathologiques négatifs, aux dépens des aspects positifs.

Bech a ainsi proposé en 1990 de mesurer la QV à l'aide d'auto-questionnaires de détresse subjective, prévus pour des névrotiques, comme le *General Health Questionnaire* (GHQ) de Goldberg ou la *Symptom Checklist* (SCL) de Derogatis. La SCL ne comporte que des items unipolaires de symptômes, alors que le GHQ comporte des items unipolaires de bien-être, à côté d'items unipolaires de symptômes. Dans son ouvrage plus récent (1993), Bech classe la SCL, le GHQ et le questionnaire de Dupuy, parmi les échelles d'inconfort ou de détresse psychologique, et les distingue des échelles de QV, parmi lesquelles il inclut l'échelle SF-36, développée par l'équipe de Wells *et al.* (1989). Cet instrument, conçu comme un instrument générique, mesure cependant plus la santé perçue que la QV.

Plus récemment l'OMS a commencé de développer des instruments propres (Bradley *et al.*, 1992), dont un grand nombre d'items provient des échelles d'anxiété et de dépression de Zung. Les items de ces échelles de Zung, comme ceux du GHQ, mesurent soit des aspects positifs, soit des symptômes, et ne sont pas bipolaires.

Évaluation de la QV des patients psychotiques

Les pathologies psychotiques délirantes chroniques (schizophrénie, délires chroniques) mais également les troubles graves de la personnalité, et les psychoses « affectives » (psychose maniaco-dépressive), que nous désignerons sous le terme de « troubles psychiatriques chroniques » (TPC), posent des questions assez spécifiques, en raison de l'importance des problèmes de réhabilitation et de réinsertion dans la communauté, et des particularités cliniques.

Dès le début des années 80, plusieurs instruments d'évaluation de la QV et/ou de l'adaptation psychosociale des TPC avaient été élaborés dans le cadre du programme de soutien dans la communauté des TPC, élaboré par le *National Institute of Mental Health* ou NIMH (Tessler et Goldman, 1982), qui avait proposé l'amélioration de la QV comme objectif principal des soins.

Les difficultés de l'évaluation de la QV subjective sont plus importantes avec les TPC qu'avec la plupart des autres troubles. En effet, les symptômes des TPC, notamment dans le cas de la schizophrénie, ont un impact important sur le mode de recueil des informations sur la QV, ainsi que sur la QV elle-même (Diener, 1984 ; Lehman et Burns, 1990).

Les symptômes « négatifs » de la schizophrénie (désintérêt, absence de motivation, troubles de la pensée et de la logique, anhédonie, etc.) peuvent ainsi déboucher sur de véritables incuries, avec négligence des besoins fondamentaux (repas, sommeil, sécurité, propreté personnelle et du cadre de vie, santé physique et mentale, etc.). Les patients se plaignent peu de ces symptômes, qui ont probablement une dimension

défensive, et soulagent la souffrance psychotique, mais retentissent parfois sur la QV du voisinage.

L'importance accordée aux domaines de vie par certains schizophrènes déficitaires peut devenir si faible (en dehors de quelques activités stéréotypées, comme fumer ou regarder la télévision) que les indices de satisfaction ou d'insatisfaction subjective individuelle n'ont plus guère de sens. Dans un travail de 1994, Chambon *et al.* attirent l'attention sur le fait que les patients schizophrènes sont souvent plus satisfaits de leur vie que nous l'imaginerions *a priori*, en raison d'une diminution de leurs attentes adaptées à leurs capacités. Ces auteurs parlent d'un effet « plafond », et soulignent que les programmes visant à améliorer les compétences sociales, risquent dans un premier temps, d'entraîner une diminution de la QV subjective de ces patients, alors que les compétences sociales augmentent, la baisse de la QV résultant plutôt de l'augmentation des attentes.

Bech (1990) a insisté quant à lui sur la capacité d'hédonisme, et différencié les patients capables d'hédonisme, névrotico-normaux, qui partagent la croyance, consciente ou non, que la quête du plaisir est un besoin humain de base, et les « anhédoniques », schizophrènes ou états-limites, qui ne partagent pas cette croyance, l'évaluation de la QV subjective étant surtout utile chez les sujets hédoniques. La présence de symptômes « positifs » (idées délirantes, hallucinations, hostilité, agressivité et négativisme) interfère également avec l'évaluation de la QV subjective par le biais des refus de participation et/ou de remplissage des questionnaires, et par des distorsions cognitives.

En raison de ces problèmes cliniques, la plupart des instruments existants sont des entretiens semi-standardisés (Lehman et Burns, 1990), et non des auto-questionnaires, bien que de nombreux patients puissent répondre à des auto-questionnaires. Les entretiens semi-standardisés permettent de recueillir les appréciations subjectives du patient, mais aussi des informations indirectes ou niées par le patient, ainsi que de consigner des manifestations comportementales. On a vu plus haut les difficultés rencontrées avec ce type d'instrument, et l'on sait qu'il peut y avoir des différences nettes d'appréciation entre soignants et patients quant à ce qui contribue le plus à la QV du patient (Thapa *et al.*, 1989).

Les principaux instruments, visant à mesurer la QV de patients souffrant de TPC, sont décrits par Lehman et Burns (1990), qui donnent des informations sur le « *Community Adjustment Form* » de Stein et Test (1980), la « *Quality of Life Checklist* » de Malm (1981), le « *Quality of Life Interview* » de Lehman (1982), la « *Satisfaction with Life Domains Scale* » de Baker et Intagliata (1982), « *l'Oregon Quality of Life Questionnaire* » de Bigelow (1982), et la « *Quality of Life Scale* » de Heinrichs *et al.* (1984). Le nombre d'items est très variable, allant de 15 (échelle de Baker et Intagliata) à 246 (échelle de Bigelow *et al.*).

Les mieux développés au plan psychométrique sont l'*Oregon Quality of Life Questionnaire* (Bigelow *et al.*, 1982) et le *Quality of Life Interview* (Lehman, 1982). Pour les autres instruments listés (SLDS, QLC, CAF), nous ne possédons aucune indication sur les coefficients de fiabilité.

Plusieurs instruments comportent des sections dites de fonctionnement, sauf l'échelle de Baker et Intagliata, uniquement centrée sur les satisfactions et qui n'explore que la QV subjective. Les satisfactions par domaine sont étudiées dans presque tous les instruments, sauf dans l'échelle de Bigelow *et al.*, et dans l'échelle de Heinrichs *et al.* (si l'on excepte un item). Cette dernière échelle, bien que nommée

« *Quality of Life Scale* » (Heinrichs, 1984) est principalement une échelle de symptômes négatifs de la schizophrénie, et de QV objective.

L'instrument de Lehman, avec celui de Malm, sont les plus complets, puisqu'ils comportent à la fois une évaluation de la QV subjective et de la QV objective, au sens des « ressources », et au sens des comportements. Les items d'insatisfaction du questionnaire de Malm ne permettent qu'une réponse oui/non.

Au regard des travaux déjà réalisés avec ces divers instruments, de leur contenu, et des éléments de validation existants, le questionnaire de Lehman paraît le plus recommandable. Une traduction de ce questionnaire a été proposée par Chambon et Trinh (1991), qui ont également traduit l'échelle de Baker et Intagliata (1982). Une vérification de la fiabilité et de la validité de l'échelle de Lehman par une étude sur le terrain est en cours (Guelfi *et al.*).

Une autre traduction de l'échelle de Baker et Intagliata a été effectuée par Mercier et Corten, pour une étude transculturelle comparant patients belges et québécois (Mercier, 1992; Corten, 1992), qui a d'ailleurs mis en évidence quelques différences culturelles d'interprétation des questions. Une analyse des correspondances multiples a été effectuée, qui montrait un premier axe dit de QV subjective, ou de bien-être, fortement influencé par l'item dépression. Le second axe opposait handicap et performance et le troisième conformisme et anticonformisme social.

Plus récemment, lors d'une étude pilote évaluant l'effet d'un programme cognitivo-comportemental « d'entraînement aux compétences sociales », chez des schizophrènes chroniques, Chambon *et al.* (1994) ont utilisé une partie de l'instrument PQVS développé par Gerin et Dazord (1991).

D'autres instruments multidimensionnels plus anciens, plutôt centrés sur les compétences ou l'adaptation sociales (QV objective) ont été proposés à une époque où le terme de QV était peu usité, mais leur contenu, ou celui de certaines de leurs sections, est parfois proche de celui des questionnaires modernes de QV, lorsqu'ils comportent des items de satisfaction subjective. (QV subjective) :

– Échelles d'adaptation sociale de Katz et Lyerly, 1963) : instrument multidimensionnel comportant à la fois une évaluation de l'adaptation sociale et de la satisfaction. Initialement destiné aux troubles névrotiques, il contient des modules permettant l'évaluation de la QV dans les TPC.

– *Social Adjustment Scale-II* de Schooler *et al.* (1979) : instrument à destination des psychotiques, issu du SAS-I (Weissman *et al.*, 1978) lequel était destiné à l'évaluation de l'adaptation sociale de patients névrotiques. Il contient des items d'adaptation (de fonctionnement) et de satisfaction.

– *Social Stress and Functioning Inventory for Psychotic Disorders* ou SSFIPD (Serban, 1978). Prévu pour l'évaluation de l'ajustement social et du stress chez les sujets psychotiques. Des items d'évaluation subjective des difficultés relationnelles le rapprochent des instruments de QV. La fiabilité test-retest du SSFIPD est médiocre (coefficients de corrélation variant de 0,43 à 0,77).

D'autres questionnaires ont été mis au point pour les besoins d'études ponctuelles, ainsi dans l'étude de Gibbons et Butler (1987) citée par Cialdella (1992), et d'Okin et Pearsall (1993); dans d'autres cas, un index global plus ou moins composite a été utilisé (Tavelli, 1990). Plus récemment, ont été proposés le questionnaire d'insatisfaction de McDonald *et al.* (1988), le *Client's Quality of Life Instrument* ou CQLI par Mulkern et Manderschied (1989) et la *Subjective Quality of Life Analysis* ou SQUALA de Zannotti et Pringuey (1992).

Ce dernier instrument, développé par une équipe française, est un autoquestionnaire portant sur les satisfactions par rapport à des domaines de vie, et sur les importances accordées aux satisfactions/insatisfactions. En ce sens, il rappelle de près le questionnaire PQVS de Gerin et Dazord (1991). Les domaines de vie sont : la santé, l'autonomie physique, le bien-être mental, l'environnement, l'habitat, le sommeil, les repas, les relations intrafamiliales et extrafamiliales, la descendance, les soucis de soi, l'amour, la sexualité, la politique, les croyances, les loisirs, la sécurité, le travail, la justice, la liberté, l'art, la beauté et l'argent.

Les données sur la validation de l'échelle SQUALA ne sont pas encore connues, en dehors d'une analyse en composantes principales.

CONCLUSIONS ET APPLICATIONS

L'évaluation de la QV est fortement conditionnée par la définition opérationnelle que l'on utilise pour qualifier ce concept. Nous avons privilégié une approche psychologique, axée sur la QV subjective, et reposant de façon prioritaire sur le concept de satisfaction, qui préjuge d'une dimension cognitive, auto-évaluative; le concept de bien-être représentant le pendant affectif de ce jugement cognitif.

Dans le domaine de la santé, on a eu trop tendance à réduire la QV à des évaluations en termes de performances ou de santé, quand il ne s'agissait pas de confort (effets secondaires), mais le principe d'étudier la QV sous son angle subjectif s'impose progressivement.

Les instruments mesurant la QV peuvent être des évaluations simples, parfois réduites à un seul item, ou bien comporter plusieurs dizaines de questions, groupées en domaines de vie, ce qui rend les processus de validation complexes.

Les instruments développés en psychiatrie, et correctement validés, sont assez peu nombreux, et attendent d'être plus utilisés. L'évaluation de la QV chez les patients névrotiques ne présente pas de particularités par rapport à celle de patients souffrant de pathologies somatiques.

Concernant les patients souffrant de TPC, on réalise habituellement les évaluations de QV subjective, à partir d'entretiens semi-structurés, sans oublier les aspects objectifs de la QV, ni l'adaptation psychosociale objective qui restent des critères pertinents de l'évaluation des dispositifs et des programmes de soins. Au plan psychométrique et pragmatique, deux instruments peuvent être recommandés chez ce type de patients, le questionnaire de Lehman (1990) et celui de Baker et Intagliata (1982), le premier étant plus exhaustif, le second plus rapide à passer.

Les applications des évaluations de QV sont nombreuses et recouvrent les recherches descriptives, les évaluations des dispositifs et programmes de soin. Au sein des programmes de soin, l'apport de l'évaluation de la QV a été envisagé à propos des essais de psychotropes, des soins de réadaptation psychosociale et des psychothérapies (Chambon et Trinh, 1991 ; Chambon *et al.*, 1994).

L'avenir du concept de QV en psychiatrie est incertain, surtout dans le domaine des essais thérapeutiques où les échelles classiques ont encore un rôle majeur en raison de leur sensibilité au changement. Concernant les évaluations plus globales de dispositifs de soins ou de programmes de réadaptation, la tendance actuelle accorde une priorité au point de vue de l'individu, mais il serait intéressant d'élargir la

perspective QV à celle de l'entourage ou du réseau social du patient, dans une approche plus écologique.

BIBLIOGRAPHIE

ABBEY A., ANDREWS F. M. — Modeling the Psychological Determinants of Life Quality. *Social Indicators Research, 16*, 1-34, 1985.

ANDREWS F. M., MCKENNEL A. C. — Measures of Self-Reported Well-Being : their Affective, Cognitive and Other Components. *Social Indicators Research, 8*, 127-155, 1980.

BAKER F., INTAGLIATA J. — Quality of Life in the Evaluation of Community Support Systems. *Evaluation and Program Planning, 5*, 69-79, 1982.

BECH P. — Quality of Life in Psychosomatic Research. In : BERNER P., GABRIEL E. (Eds), *Psychopathology n°20.* Karger, Basel, 169-179, 1987.

BECH P. — Methodological Problems in Assessing Quality of Life as Outcome in Psychopharmacology : a Multiaxial Approach. In : BENKERT O., MAIER W., RICKELS K. (Eds), *Methodology and Evaluation of Psychotropic Drugs.* Springer-Verlag, Berlin, Heidelberg, 121-129, 1990.

BECH P. — *Rating Scales for Psychopathology, Health Status, and Quality of Life : a Compendium on Documentation in Accordance with the DSM-III-R and WHO Systems.* Springer-Verlag, Berlin, Heidelberg, 1993.

BIGELOW D. A., BRODSKY G., STEWARD L., OLSON M. — The Concept and Measurement of QOL as a Dependent Variable in Evaluation of Mental Health Services. In : STAHLER G. J., TASH W. R. (Eds), Innovative Approaches to Mental Health Evaluation, Academic Press, New York, 345-366, 1982.

BRADLEY C., MEADOWS K. A., SOWDEN A. J. — *General Well-Being and Satisfaction with Treatment Scales for Use with People with Diabetes Requiring Insulin. Part. 1 : Psychometric Development and Retranslation of the English, French, and German Versions.* Report of the World Health Organisation, Regional Office for Europe, Copenhagen, June 1992.

CAMPBELL A., CONVERSE P. E., RODGERS W. L. — *The Quality of American Life.* Russel Sage Foundation, New York, 1976.

CHAMBON O., TRINH A. — L'intérêt de l'évaluation de la qualité de vie des malades mentaux chroniques lors du processus de réadaptation sociale. *Synapse (Fr), n° 77*, 77-86, 1991.

CHAMBON O., MARIE-CARDINE M., DAZORD A. — *Rehabilitation and Psychotherapy of Schizophrenic Patients : Toward a Cognitive Approach in France. Result of a Preliminary Study on Social Skills Training and Quality of Life.* European Psychiatry, special issue, (in press).

CIALDELLA Ph. — Évaluation de la qualité de vie des patients psychiatriques chroniques vivant dans la communauté. In : R. LAUNOIS et F. RÉGNIER (Eds), *Décision thérapeutique et qualité de vie*, John Libbey Eurotext, Paris, 113-126, 1992.

CORTEN Ph. — *La qualité subjective de la vie : concept d'évaluation en clinique de la réadaptation.* Communication au colloque qualité de vie et santé mentale : aspects conceptuels et méthodologiques. Lyon, 9-11 décembre 1992.

DAZORD A., GERIN P., BROCHIER C., CLUSE M., TERRA J. L., SEULIN C. — Un modèle de qualité de la vie subjective adapté aux essais thérapeutiques : intérêt chez les patients dépressifs. *Santé Mentale au Québec, 18*, 2, 49-74, 1993.

DIENER E. — Subjective Well-Being. *Psychological Bulletin, 95*, 3, 542-575, 1984.

DUNBAR G. C., STOKER M. J., HODGES T. C. P. — *The Development of Beecham's Quality of Life Scale.* Communication au congrès mondial de Psychiatrie, Athènes, 1989.

ENDICOTT J., NEE J., HARRISON W., BLUMENTHAL R. — Quality of Life Enjoyment and Satisfaction Questionnaire : a New Measure. *Psychopharmacol. Bull., 29*, 321-326, 1993.

FEIGHNER J. P. — Impact of Anti-Anxiety Therapy on Patient's Quality of Life. *The American Journal of Medicine, 82*, suppl. 5A, 14-19, 1987.

GERIN P., DAZORD A., CIALDELLA Ph., LEIZOROVICZ A., BOISSEL J. P., et le groupe de recherche OCAPI — Le questionnaire Profil de la qualité de vie subjective (PQVS). *Thérapie, 46*, 131-138, 1991.

GLAZER W. M., AARONSON H. S., PRUSOFF B. A., WILLIAMS D. H. — Assessment of Social Adjustment in Chronic Ambulatory Schizophrenics. *Journal of Nervous and Mental Disease, 168*, 8, 493-497, 1980.

HEINRICHS D. W., HANLON T. E., CARPENTER W. T. Jr — The Quality of Life Scale : an Instrument for Rating the Schizophrenic Deficit Syndrome. *Schizophrenia Bulletin, 10*, 3, 388-398, 1984.

JOYCE C. R. B. — Quality of Life : the State of the Art in Clinical Assessment. In : WALKER S. R., ROSSER R. M., (Eds), *Quality of life : Assessments and Application.* MTR Press, Lancaster, 169-179, 1987.

LEHMAN A. F., WARD N. C., LINN L. S. — Chronic Mental Patients : the Quality of Life Issue. *Am. J. Psychiatry, 10*, 1271-1276, 1982.

LEHMAN A. F. — A Quality of Life Interview for the Chronically Mentally Ill. *Evaluation and Program Planning, 11*, 51-62, 1988.

LEHMAN A. F., BURNS B. — Severe Mental Illness in the Community. In : SPILKER B. (ed), *Quality of Life Assessments in Clinical Trials*, Raven Press, New York, 357-366, 1990.

MACDONALD L., SIBBALD B., HOARE C. — Measuring Patient Satisfaction with Life in a Long-Stay Psychiatric Hospital. *The International Journal of Social Psychiatry, 34*, 292-304, 1988.

MALM U., RAY P. R. A., DENCKER S. J. — Evaluation of the Quality of Life of the Schizophrenic Patient : a Checklist. *Schizophrenia Bulletin, 7*, 3, 477-487, 1981.

MARKOWITZ J. S., WEISSMAN M. M., OUELLETTE R., LISH J. D., KLERMAN G. L. — Quality of Life in Panic Disorder. *Arch. Gen. Psychiatry, 46*, 984-992, 1989.

MASLOW A. — A Theory of Human Motivation. *Psychological Review, 50*, 370-396, 1943.

MERCIER C. — *La méthodologie de l'évaluation de la qualité de vie subjective en psychiatrie.* Communication au colloque qualité de vie et santé mentale : aspects conceptuels et méthodologiques. Lyon, 9-11 décembre 1992.

MULKERN V. M., MANDERSCHIED R. W. — Characteristics of Community Support Program Clients in 1980 and 1984. *Hospital and Community Psychiatry, 40*, 165-172, 1989.

OKIN R. L., PEARSALL D. — Patient's Perceptions of their Quality of Life 11 Years after Discharge from a State Hospital. *Hospital and Community Psychiatry, 44*, 3, 236-240, 1993.

PERAUT P. — *Qualité de vie : aspects historiques et sociologiques, application en médecine et psychiatrie.* Communication au colloque qualité de vie et santé mentale : aspects conceptuels et méthodologiques. Lyon, 9-11 décembre 1992.

SERBAN G. — Social Stress and Functioning Inventory for Psychotic Disorders (SSFIPD) : Measurement and Prediction of Schizophrenics' Community Adjustment. *Comprehensive Psychiatry, 19*, 4, 337-346, 1978.

SCHOOLER N., HOGARTHY G., WEISSMAN M. — Adjustment Scale-II (SAS II). In : HARGREAVES W. A., ATTKISSON C. C, SORENSON J. E. (Eds), *Resource Materials for Community Mental Health Programs Evaluators.* US department of Health, Education and Welfare, Rockville, 290-303, 1979.

SHUMAKER S. A., ANDERSON R. T., CZAJKOWSKI — Psychological Tests and Scales. In : SPILKER B. (ed), *Quality of Life Assessments in Clinical Trials*, Raven Press, New York, 1990.

SIMPSON C. J., HYDE C. E., FARAGHER E. B. — The Chronically Mentally Ill in Community Facilities : a Study of Quality of Life. *Br. J. Psychiatry, 154*, 77-82, 1989.

STEIN L. I., TEST M. A. — Alternative to Mental Health Hospital Treatment : I. Conceptual Model, Treatment Program and Clinical Evaluation. *Arch. Gen. Psychiatry, 37*, 392-397, 1980.

TESSLER R. C, GOLDMAN H. H. — *The Chronically Mentally Ill: Assessing Community Support Programs.* Ballinger, Cambridge (Mass), 1982.

TEST M. A., STEIN L. I. — Alternative to Mental Hospital Treatment : III Social Cost. *Arch. Gen. Psychiatry, 37*, 409-412, 1980.

THAPA K., ROWLAND L. A. — Quality of Life Perspectives in Long-Term Care : Staff and Patient Perceptions. *Acta Psychiatr. Scand., 80*, 267-271, 1989.

WEISSMAN M. M., PRUSOFF B. A., THOMSON W. D. *et al.* — Social Adjustment by Self-Report in a Community Sample and in Psychiatric Outpatients. *Journal of Nervous Mental Disease, 166*, 317-326, 1978.

WEINSTEIN M. C. — Décisions médicales et politique de santé. Le nécessaire compromis entre des objectifs contradictoires : qualité de vie, quantité de vie, et coût. In : R. LAUNOIS et F. RÉGNIER, *Décision thérapeutique et qualité de vie.* John Libbey Eurotext (Eds), Paris, 65-78, 1992.

WELLS K. B., STEWART A., HAYS R. D., BURNAM A., ROGERS W., DANIELS M., BERRY S., GREENFIELD S., WARE J. — The Functioning and Well-Being of Depressed Patients : Results from the Medical Outcomes Study. *JAMA, 262*, 7, 914-919, 1989.

ZAUTRA A. J., GOODHART D. — Quality of Life Indicators : a Review of the Literature. *Community Mental Health Review, 4*, 1, 1-10, 1979.

ZANNOTTI M., PRINGUEY D. — *Construction et utilisation d'une échelle de qualité de vie : la SQUALA*. Communication au colloque qualité de vie et santé mentale : aspects conceptuels et méthodologiques. Lyon, 9-11 décembre 1992.

18

ÉVALUATION DES PSYCHOTHÉRAPIES

par A. DAZORD

INTRODUCTION

L'évaluation des effets des psychothérapies à l'aide d'enquêtes et d'outils statistiques est une discipline relativement récente. Développée dans les pays anglo-saxons depuis quelques 40 ans (Garfield et Bergin, 1986; Luborsky, 1988), ce n'est que dans les années 80 que ce type de recherche a été introduit en France, en particulier par notre équipe. Le contexte psychiatrique français de l'époque, marqué à la fois par la psychanalyse et les recherches biopharmacologiques de pointe, a donné une spécificité un peu particulière à notre démarche. Nous avons de manière constante cherché d'une part à anticiper les critiques des psychanalystes à propos du risque de réification et de réductionnisme, tandis que d'autre part nous nous efforcions de définir des stratégies méthodologiques fiables et rigoureuses, en accord avec les exigences des fondamentalistes.

Nous consacrerons ce chapitre essentiellement aux apports de la recherche française, renvoyant le lecteur pour ce qui est des travaux anglo-saxons aux ouvrages de référence en langue anglaise (Garfield et Bergin, 1986; Luborsky, 1988) ou aux traductions françaises effectuées par notre équipe (Gerin et Dazord, 1992). Quant aux études qui seront présentées (Dazord *et al.* ou Gerin *et al.*, 1991, 1992, 1993), soulignons qu'elles ont été essentiellement conduites dans des institutions. Elles sont mentionnées ici davantage pour montrer le type d'étude qui peut être réalisé, que pour susciter une réflexion sur les diverses approches psychothérapiques envisagées.

SPÉCIFICITÉ DE LA RECHERCHE SUR LES PSYCHOTHÉRAPIES

La modélisation du questionnement clinique

De manière plus impérative que dans le cadre de la psychopathologie dite « quantitative », pour ce qui est de l'évaluation des psychothérapies, un des principaux objectifs est d'arriver à modéliser le questionnement spécifique lié *au processus psychothérapique*. Cette étape doit se dégager des modèles médicaux classiques centrés sur les *symptômes* ou la *nosographie* ou de la *psychométrie* envisagée sous l'angle des « tests mentaux » ou des « tests de personnalité ». Certes, ces échelles classiques ont leur intérêt pour définir les groupes de sujets sur lesquels les études sont pratiquées, ou pour suivre leur évolution, mais l'objectif des recherches évaluatives

sur les psychothérapies est autre : il s'agit essentiellement de soumettre les hypothèses cliniques au crible d'analyses rigoureuses portant sur des observations répétées.

Un des chercheurs exemplaires à cet égard est L. Luborsky (Luborsky, 1988). Plutôt qu'utiliser les listes-symptômes classiques issues de la nosographie psychiatrique, il a construit, pour évaluer l'état des patients, une échelle (dite « échelle Santé-Maladie ») dont le principe de base consiste à faire appel à un mode de fonctionnement clinique spontané et habituel, à savoir la comparaison entre patients. L'échelle comporte ainsi une trentaine de cas-type classés par gravité croissante; il est demandé de chercher, dans cette liste de cas-type, celui qui se rapproche le plus du patient à évaluer et d'attribuer alors à ce dernier la note correspondante. Dans un registre différent, Luborsky a cherché à rendre compte de la « conflictualité fantasmatique » des sujets. Partant d'enregistrements de séances, il repère dans des fragments de discours du sujet des « épisodes relationnels » le mettant en scène avec une autre personne y compris le thérapeute ou lui-même. Des lectures répétées de ces épisodes ont alors pour but de se focaliser sur les souhaits (ou désirs) et les réponses à ces souhaits. On constate en général une très forte redondance du discours : un souhait donné revient de manière prédominante au cours d'une série d'épisodes relationnels. Le thème conflictuel relationnel central est constitué par ce souhait et les réponses qui l'accompagnent. Les variations des réponses permettent de suivre ainsi les effets de psychothérapies à un niveau du processus qui n'est pas nécessairement conscient.

Les stratégies

Une problématique satisfaisante et des instruments adéquats

Deux conditions sont indispensables pour mener à bien ce type d'études : disposer d'instruments capables de prendre en compte les éléments déterminants intervenant dans une psychothérapie, et qui sont dans le registre du subjectif, mais parallèlement être à l'écoute des cliniciens.

• *Prendre en compte des éléments subjectifs, et en particulier le discours des différents partenaires, qui rendra compte de leurs représentations*

Il est non seulement nécessaire de disposer d'instruments, questionnaires le plus souvent, explorant des paramètres subjectifs, ou psychodynamiques tels les motivations du patient, du soignant, son degré d'*insight*, la qualité de sa relation avec le soignant... De tels instruments existent, ont fait l'objet d'études de validation. Nous en avons présenté une revue dans un ouvrage récent (Gerin, 1991).

Il nous est surtout apparu primordial de pouvoir rendre compte des *représentations* qu'avaient les différents partenaires de la scène clinique (patient, thérapeute,...) des difficultés ressenties tant dans la vie du patient que lors du déroulement de la psychothérapie. À cet effet, nous avons, à partir de discours écrits (obtenus en réponse le plus souvent à des questions simples du type « *qu'est-ce qui ne va pas ?* »), adapté les méthodes d'analyse de contenu classique. Il en est résulté l'obtention de grilles de lecture, permettant de dégager, à partir de ce discours, de grandes catégories (nature des difficultés, manière d'en parler...). Ces catégories sont alors affectées d'un code qui permet d'en faire une analyse statistique. À partir de ces catégories codées, il a été possible de calculer des « indices » un peu plus complexes, évaluant

par exemple le degré de mobilité d'un patient (variation de ces dites catégories), son degré d'ouverture (nombre de catégories différentes), son degré d'accord ou de désaccord avec divers intervenants (nombre de catégories communes mentionnées) (Dazord, 1992, 1993). Ce type d'instrument, là encore, peut se valider; il suffit d'évaluer la concordance ou les corrélations existant avec des réponses à d'autres questionnaires, ou éventuellement de comparer, dans certains cas, les réponses faites par des «juges» différents (thérapeutes) avec celles des patients (Gerin, 1991).

En ce qui concerne le traitement statistique de ce type de données, soulignons qu'on a soit à faire à des catégories, sur lesquelles on peut effectuer des comptages, ou soit à des mesures d'ordre, où 4 est simplement plus grand que 2, mais non le double comme dans les mesures d'intervalles. Il existe des tests statistiques appropriés à ce type de mesure ou tests non paramétriques.

• *Prendre en compte la complexité de la scène clinique et le point de vue de différents partenaires*

Il ne suffit pas d'observer les variables supposées avoir le plus d'importance d'après le modèle conceptuel propre aux cliniciens concernés par une enquête donnée. Encore faut-il prendre aussi en considération celles envisagées dans d'autres modèles conceptuels, afin d'éviter de suivre une démarche tautologique, une démarche qui ne pourrait que confirmer le rôle des facteurs présupposés car elle ne donnerait pas à d'autres facteurs la possibilité d'être observés. L'importance de respecter la complexité de la situation clinique implique l'utilisation de profils, décrivant par exemple le patient à la fois sur le plan de ses symptômes, de ses centres d'intérêt, de ses attitudes, de sa qualité de vie. Plus que sur un indice unique, c'est dans un plan qu'il faut travailler et même dans plusieurs plans, car il peut être intéressant de confronter les points de vue du patient, du thérapeute, d'autres intervenants, de la famille. Là encore, soulignons que cette complexité peut faire l'objet d'une analyse statistique. Un certain nombre de précautions doivent être prises lorsque des tests multiples sont pratiqués, et une plus grande exigence s'impose quant au seuil de significativité à utiliser (Glantz, 1981). Par ailleurs, il existe, en dehors de la statistique classique qui traite d'un nombre restreint de caractères sur un petit ensemble d'individus, avec des hypothèses probabilistes très restrictives, d'autres méthodes dites d'analyse des données, consistant en des études plus globales, des individus et des variables les caractérisant, avec des représentations graphiques suggestives. Le principe consiste à représenter les proximités entre individus ou entre leurs réponses, sur un graphique.

• *Travailler sur les hypothèses des cliniciens*

Tout au long des dix années pendant lesquelles nous avons conduit des projets de recherche avec des équipes cliniques, il nous est apparu que cet enracinement dans la pratique clinique donnait naissance à des hypothèses heuristiques. Nous pouvons en donner deux types d'exemples.

Au cours de la mise au point d'un protocole d'étude sur le processus psychothérapeutique lors d'interventions de crise, dans un centre genevois (Dazord, 1991, 1992) a surgi l'hypothèse qu'un accord trop important entre le patient et son thérapeute risquait d'avoir pour corollaire une absence de mobilisation du patient; mais si le désaccord était trop important, l'interaction ne pouvait avoir lieu : le patient et le

thérapeute étant sur deux planètes différentes. Le protocole élaboré en fonction de cette hypothèse a comporté l'étude des représentations des deux partenaires à propos des difficultés du patient (cf. *supra* p. 185-186). Les résultats trouvés ont permis d'affiner l'hypothèse initiale, mais aussi d'en suggérer de nouvelles : parmi les facteurs semblant influer sur le changement à long terme des patients, il apparaissait en effet que leur degré d'ouverture ou de focalisation (concernant les difficultés du patient) à différents moments du soin, de même que leur mobilité concernant la manière d'envisager les difficultés en cause, étaient particulièrement importants à considérer.

Dans un hôpital de jour (Seulin, 1989), les soignants cherchaient à évaluer l'impact de la relation soignant-soigné sur l'évolution des patients. Les échelles disponibles, ne prenant pour la plupart en compte que l'absence d'investissement des patients, ou des degrés d'investissements croissants, leur apparurent négliger les aspects liés à la disqualification du soin. Les thérapeutes procédèrent alors selon le principe qui a présidé à la mise sur pied de l'échelle-santé maladie de Luborsky (*cf.* p. 185) : à partir du recueil de cas cliniques évoquant, pour les thérapeutes concernés par la recherche, divers degrés d'« engagement » dans le processus, une échelle fut construite. Son utilisation dans une enquête ultérieure (*cf. supra* p. 186) a permis de réfléchir sur l'impact conjugué dans l'évolution des patients, de leur engagement, de leur type clinique et de la spécificité des diverses approches psychothérapeutiques.

Une méthodologie rigoureuse

• *La validation des instruments*

Elle n'est pas spécifique à l'évaluation des psychothérapies. Elle obéit à une stratégie codifiée (Carmines et Zeller, 1979), qui consiste à évaluer, dans un questionnaire, avec des outils statistiques adéquats (coefficients de corrélation, de concordance,...) la cohérence entre les réponses obtenues à des questions semblables, la cohérence entre les réponses obtenues à des questionnaires voisins, entre les points de vue des différents intervenants et ce à des temps différents. Nous avons particulièrement été frappés par le fait que les sujets, même alcooliques (qui sont habituellement exclus des protocoles d'évaluation thérapeutique) ou psychotiques, ne répondent pas au hasard.

• *Se donner les moyens de travailler sur des sous-groupes*

Nous avons très vite constaté que nombre de questions ne trouvaient aucune réponse lorsque l'échantillon des patients était trop hétérogène. Vouloir par exemple comparer différentes approches psychothérapiques sur un groupe hétérogène de patients hospitalisés dans un service donné n'aboutit en général à aucun résultat significatif et d'ailleurs n'a pas grand sens. En revanche, l'utilisation d'échelles diagnostiques (CIM-10, DSM-III-R, ou même diagnostic utilisé couramment par les soignants) permet de discriminer les différentes approches entre elles. D'une manière générale, nous avons pu montrer, dans un certain nombre d'études, que les approches psychothérapeutiques en relation avec l'évolution des patients n'étaient pas les mêmes chez les patients psychotiques ou chez les névrotiques, et que les profils de changements obtenus différaient également (Gerin et Dazord, 1992).

• *Les protocoles de recherche*

D'une manière générale, il existe deux grands types d'études évaluatives. Le premier type d'études vise une description plus fine de la situation observée ou la proposition d'un modèle explicatif. Le second a pour but de vérifier un modèle explicatif antérieurement proposé, et ses possibilités de généralisation. La première approche, heuristique, dégage une hypothèse explicative à partir de l'observation. C'est la plus courante. La seconde, inférentielle, vérifie la validité générale de telles hypothèses. C'est l'approche utilisée dans les essais thérapeutiques. Elle implique d'éliminer entre autres tous les biais de sélection, d'utiliser une répartition aléatoire des patients, des techniques en simple ou double aveugle. Elle n'est qu'exceptionnellement réalisée dans le champ des approches psychothérapeutiques.

• *Les outils statistiques*

Nous avons mentionné plus haut l'intérêt des tests non paramétriques. Il faut également savoir qu'il existe à côté des statistiques classiques, des méthodes très adaptées aux données qualitatives et aux approches non inférentielles qui sont les méthodes d'analyses des données et qui, le plus souvent, utilisent des représentations graphiques suggestives où les individus et leurs réponses se projettent en fonction de leur proximité (Volle, 1985).

LES RÉSULTATS OBTENUS

Les effets des psychothérapies

D'une manière générale, l'ensemble de nos résultats a eu pour intérêt de montrer, comme d'ailleurs les études anglo-saxonnes, que dans la plupart des cas les prises en charges psychothérapeutiques s'accompagnaient d'une évolution favorable. Il est difficile néanmoins d'en donner une appréciation quantitative simple. Certains ont proposé d'évaluer les changements sous la forme d'un indice dit « taille d'effet » qui norme les résultats en fonction de la dispersion de la mesure. Des méta-analyses objectiveraient ainsi une « taille d'effet » voisine de 1 pour les approches psychothérapeutiques ; par comparaison, les tailles d'effets obtenues avec les médications antidépressives varieraient de 0,4 à 0,8 (Lambert, Shapiro et Bergin, 1986).

Nous avons, dans la plupart de nos études, confirmé ce type de résultats, mais il nous est surtout paru intéressant de voir quels types de patients changeaient, et sur quels plans. À titre d'exemple, nous citerons une étude effectuée auprès de sujets ayant bénéficié de psychothérapies (de type analytique pour la plupart) dans un secteur psychiatrique public et universitaire genevois (Aapro *et al.*, 1994). Les patients se présentaient cliniquement comme dépressifs et anxieux, avec une immaturité relationnelle et une pauvreté des relations interpersonnelles non négligeables. Après un an de prise en charge, la dépression s'est améliorée de manière notable, l'anxiété à un degré moindre, mais l'immaturité relationnelle est demeurée inchangée.

Les facteurs de changement

L'étude des facteurs de changement se fait essentiellement en regardant les relations ou corrélations entre l'état clinique des patients après la prise en charge psychothérapeutique et un certain nombre de caractéristiques initiales du patient, ou de la prise en charge effectuée.

D'une manière générale, les facteurs de changement mis en évidence varient selon les contextes et les types de patients; il peut s'agir de la gravité symptomatique initiale, de la force du moi, d'un certain nombre d'éléments psychodynamiques (*insight*, motivations, éléments contretransférentiels), de la qualité de la relation thérapeutique, de divers aspects du processus psychothérapeutique, et de diverses caractéristiques de la prise en charge. C'est par exemple le cas de ce centre de crise genevois (Dazord, 1992), où si l'on résume les résultats trouvés, les facteurs corrélés significativement avec les changements sont : les motivations initiales du patient, la force du moi, et le processus psychothérapeutique tel qu'il est exploré par certains indices évaluant la qualité de l'intervention de crise. En revanche, l'état clinique initial est sans relation apparente avec l'évolution. Dans le même centre de crise, le recueil des représentations des soignants et des patients sur les difficultés des patients nous a permis de montrer que le fait pour le patient d'évoquer à l'arrivée dans le centre de crise un large spectre de difficultés était un élément favorable. Après la prise en charge, la focalisation du patient sur un nombre restreint de difficultés était très corrélée à une évolution ultérieure favorable, de même que la mobilité de son point de vue à propos de ses difficultés. Dans tel autre lieu (hôpital psychiatrique de la région lyonnaise), on s'aperçoit, au contraire, que c'est la gravité symptomatique initiale du patient qui est en relation étroite avec les résultats (Gerin, 1992b), ou encore dans une unité effectuant des prises en charge de type ergothérapique, les aptitudes initiales du patient dans le travail semblent prédire sa qualité de vie au bout de trois mois de prise en charge (estimée par un questionnaire que nous avons mis au point, le « profil de la qualité de vie subjective » ou PQVS (Gerin, 1992c) et un autre plus classique, le questionnaire de Baker-Intagliata) (tableau 1). À ce stade de l'exploitation des résultats, on se trouve face à une série de corrélations et donc face à une série de facteurs semblant intervenir dans les changements. Une étape supplémentaire s'impose afin de savoir si tous ces facteurs mis en évidence sont à mettre sur le même plan. Il existe des techniques statistiques (régression multiple) qui permettent de classer ces différents facteurs, d'en estimer le poids respectif. Dans le centre de crise genevois mentionné plus haut (Dazord, 1991), les facteurs les plus en relation avec l'évolution sont des éléments psychodynamiques (confiance dans le traitement, mobilisation du patient en relation avec l'intervention de crise, et coopération) tandis que dans un autre service d'hospitalisation, ce qui vient en tête c'est l'intensité de la scène clinique psychotique (Gerin, 1992b).

Les effets des différentes psychothérapies : discussion du paradoxe de l'équivalence

Les résultats anglo-saxons

Outre Atlantique, les recherches n'ont pu mettre en évidence des différences « d'efficacité » entre les diverses formes de psychothérapies, ce qui a été désigné sous

Tableau 1. *Corrélations entre les aptitudes initiales du patient et sa qualité de vie à trois mois*

	Questionnaire Baker-Intagliata				
Aptitudes au travail	*Satisf. de soi*	*Estime de soi 1*	*Estime de soi 2*	*Activités quotid.*	*Situation économique*
Présentation vestimentaire	0,03	0,17	0,05	0,02	- 0,17
Relation responsables	0,15	0,14	- 0,14	0,02	- 0,47*
Relation collègues	- 0,21	- 0,02	- 0,02	- 0,07	- 0,42*
Travail équipe	0,14	0,25	0,41*	0,18	0,28
Dextérité	0,40*	0,41*	0,35*	0,35*	0,13
Soin	0,29*	0,40*	0,33*	0,43*	0,17
Rythme	0,44*	0,40*	0,46*	0,52*	0,24
Capacités physiques	0,34*	0,29*	0,32*	0,44*	0,24
Utilisation machines	0,27	0,30	0,39	0,45*	0,31
Adaptation	0,06	0,14	0,20	0,35*	0,27
Compréhension	0,24	0,33*	0,30*	0,33*	0,31*
Nouveau problème	0,26	0,37*	0,36*	0,34*	0,32
Enchaînement logique	0,27	0,23	0,34*	0,33*	0,34*
Motivation	0,22	0,27	0,28	0,40*	0,38*
Ponctualité	0,07	0,29*	0,02	0,17	0,27
Stabilité	0,05	0,28	0,21	0,20	0,35*

	Questionnaire PQVS				
Aptitudes au travail	*Gestes usuels*	*Forme*	*Argent*	*Loisirs*	*Santé*
Présentation vestimentaire	- 0,06	0,04	- 0,05	0,06	- 0,01
Relation responsables	0,05	0,02	- 0,16	- 0,21	0,09
Relation collègues	0,27	0,08	- 0,26	- 0,06	- 0,17
Travail équipe	0,32*	0,34*	- 0,33*	0,23	0,14
Dextérité	0,30*	0,28	- 0,13	0,37*	0,32*
Soin	0,18	0,19	- 0,10	0,25	0,33
Rythme	0,39*	0,34*	- 0,09	0,43*	0,39*
Capacités physiques	0,34*	0,33*	- 0,21	0,35*	0,27
Utilisation machines	0,29	0,19	- 0,14	0,37*	0,27
Adaptation	- 0,17	- 0,19	- 0,03	0,23	0,09
Compréhension	0,13	0,18	- 0,14	0,22	0,27
Nouveau problème	0,08	0,08	- 0,13	0,19	0,20
Enchaînement logique	0,13	0,19	- 0,22	0,22	0,23
Motivation	0,10	0,02	- 0,10	0,18	0,16
Ponctualité	0,04	0,01	0,12	0,02	0,01
Stabilité	- 0,02	0,03	- 0,02	- 0,07	0,01

* = $p < 0,01$

le nom de *paradoxe de l'équivalence* (Stiles, 1986), et a conduit à rechercher d'autres éléments en relation avec les résultats. Deux facteurs principaux se dégagent :

a) de manière très constante, la qualité de la relation thérapeutique telle que la vit le patient est corrélée significativement avec une évolution favorable (Horvath et Symonds, 1991) ;

b) le deuxième facteur relié de manière régulière à la qualité clinique est le niveau général du fonctionnement mental des patients dont l'aspect positif est reflété par la « force du moi », et l'aspect négatif par le degré de dysfonctionnement (Orlinsly, 1991).

Ainsi, de ces études se dégage l'idée qu'une psychothérapie efficace est clairement l'apanage des patients prêts à y participer et dans une attitude de collaboration.

En France

Les résultats que nous avons obtenus diffèrent quelque peu. La simple prise en compte du fait de bénéficier de tel type de psychothérapie ne permet pas de différencier l'évolution des patients. Ceci ne devient possible qu'à condition de travailler sur des sous-groupes relativement « homogènes » de patients, et sur des profils de changement. Ainsi, dans l'hôpital de jour où a été construite l'échelle d'engagement mentionnée plus haut (p. 187), quatre sous-groupes de patients ont pu être individualisés à l'aide de la CIM-10, et il est particulièrement intéressant de noter que, pour les schizophrènes, ce qui est corrélé avec une bonne évolution, c'est surtout leur engagement face à l'institution; pour les psychoses affectives et dépressives, c'est davantage ce qui se passe dans les entretiens individuels et les groupes de discussion, pour les patients présentant des troubles de la personnalité, l'engagement dans les thérapies groupales (forum) est un élément prédicteur favorable alors que l'inverse est observé pour les patients névrotiques (tableau 2). Ces résultats confirment d'autres enquêtes faites à l'hôpital psychiatrique où nous avons montré qu'il était capital de ne pas se contenter de comparer les différentes approches, mais d'évaluer l'investissement des patients sur ces différents approches (Gerin, 1992b) ; de plus, il fallait prendre en compte, non seulement l'aspect formel et cadré du soin, mais aussi tous les aspects informels. Déjà, à l'époque, il nous était alors apparu que, si pour les dépressifs l'aspect bien cadré du soin était important, en revanche les patients psychotiques avaient tendance à davantage bénéficier des prises en charge informelles, et de ce qui se passait dans les couloirs de l'institution (effet « interstice » décrit par Roussillon). Dans un ordre d'idée voisin, pour recherche l'impact éventuel de facteurs spécifiques caractérisant une approche psychothérapique, il faut éviter de travailler sur des échantillons trop disparates. Ainsi, une première enquête, où étaient étudiés sur un éventail assez large de patients les effets de la musicothérapie, montrait qu'aucun des aspects spécifiques de l'approche n'était en relation avec leur évolution, seules les motivations globales des patients semblaient prédictrices (Gerin, 1992a). À l'inverse, une étude plus récente (mentionnée plus haut, p. 189), et s'adressant à un échantillon moins contrasté de patients suivant une approche « ergothérapique », montre que les éléments les plus corrélés avec l'évolution des patients et en particulier leur qualité de vie, sont leurs aptitudes initiales au travail (tableau 1). On peut donc mettre en évidence des effets spécifiques des psychothérapies à condition d'éviter de travailler sur des populations de patients trop composites, et à condition de prendre des éléments pertinents concernant ces psychothérapies, et les changements des patients. Un indice global de changement manque en général de

Tableau 2. *Corrélations entre l'engagement initial dans différentes approches psychothérapeutiques et l'évolution des patients (sortie de l'hôpital de jour)*

Troubles de la personnalité

Changements observés à la sortie	*E I*	*F O*	*D I*	*R C*	*Insti.*
Gravité globale	- 0,13	**0,65**	0,03	0,02	0,14
Autonomie	- 0,08	**0,65**	0,03	- 0,11	0,09
Sévér. sympt.	- 0,20	**0,69**	0,24	- 0,02	0,17
Malaise int.	- 0,27	**0,75**	0,17	0,03	0,15
Retent. ent.	- 0,21	**0,62**	0,14	0,00	- 0,03
Capacités	- 0,15	**0,60**	0,16	0,13	0,31
Qualité rel.	- 0,37	**0,53**	0,15	- 0,13	- 0,06
Centres int.	- 0,22	0,34	- 0,31	04	- 0,10

Troubles névrotiques

ESM à la sortie	*F I*	*F O*	*D I*	*R C*	*Insti.*	*Cr*
Gravité globale	0,08	**- 0,76**	- 0,19	- 0,58	- 0,07	**0,71**
Autonomie	0,18	**- 0,61**	- 0,21	**- 0,68**	- 0,18	0,48
Sévér. sympt.	0,29	**- 0,72**	- 0,35	- 0,30	- 0,00	**0,81**
Malaise int.	0,44	**- 0,51**	- 0,47	- 0,10	- 0,10	**0,77**
Retent. ent.	- 0,16	**- 0,71**	- 0,14	- 0,38	- 0,10	0,61
Capacités	0,11	**- 0,75**	- 0,50	- 0,47	- 0,18	0,57
Qualité rel.	- 0,02	**- 0,65**	- 0,18	- 0,46	- 0,11	0,61
Centres int.	0,08	**- 0,72**	- 0,17	**- 0,72**	- 0,04	0,60

Schizophrénies

Changements observés à la sortie	*E I*	*F O*	*D I*	*R C*	*Insti.*
Gravité globale	**0,40**	**0,50**	**0,55**	**0,49**	**0,66**
Autonomie	0,34	0,39	**0,50**	**0,51**	**0,65**
Sévér. sympt.	0,37	**0,43**	0,38	0,41	**0,57**
Malaise int.	0,23	0,18	0,45	0,29	**0,60**
Retent. ent.	**0,14**	0,36	0,43	**0,49**	**0,54**
Capacités	0,29	0,26	**0,39**	0,37	**0,72**
Qualité rel.	0,02	0,06	0,16	0,38	**0,59**
Centres int.	**0,49**	**0,47**	0,29	**0,54**	**0,77**

Légende : E I : entretiens individuels, F O : forum ; D I : groupes de discussion : R C : rencontres de couloir ; Insti : institution, Cr : ateliers de « créativité » ; **caractère gras** : p < ou = 0,05

sensibilité; par ailleurs, il peut être plus adéquat d'étudier l'évolution de la symptomatologie dépressive chez des dépressifs et celle de la symptomatologie psychotique chez des schizophrènes. L'utilisation de profils de changements à condition de respecter les précautions méthodologiques mentionnées plus haut est le plus souvent une solution satisfaisante.

Les éléments explicatifs

Ces études évaluatives concernant les psychothérapies ne se limitent pas aux seules études des facteurs de changements. Il est possible par exemple d'étudier les éléments symptomatiques en relation avec les attentes des patients face à la psychothérapie et au thérapeute; le tableau 3 (recherche effectuée dans le centre genevois mentionné plus haut, Aapro *et al.*, 1994) montre que ces attentes sont d'autant plus importantes que l'anxiété est grande et que la force du moi est faible.

Tableau 3. *Exemple de facteurs explicatifs. Psychothérapies analytiques (Genève) – 67 patients*

Relation entre la symptomatologie initiale des patients et leurs attentes face à la psychothérapie			
Symptôme	*Attente psychothérapie*	*Coeff. cor.*	*P*
Anxiété →	Amélioration symptômes	0,33	0,03
→	Comprendre	0,39	0,01
→	Performances	0,36	0,02
Difficultés sexuelles →	Origine des difficultés	0,3	0,04
→	S'exprimer	0,41	0,01
→	Amélioration symptômes	0,33	0,03
→	Comprendre	0,3	0,04

Relation entre la symptomatologie initiale des patients et leurs attentes face au thérapeute			
Symptôme	*Attente thérapeute*	*Coeff. cor.*	*P*
Anxiété →	Personnalité qui convienne	0,43	0,06
→	Conseille	0,32	0,04
→	Comprenne	0,45	0,04
→	Soutienne	0,36	0,02
Comportement d'échec →	Explique	0,32	0,05
→	Comprenne	0,33	0,04
→	Soutienne	0,38	0,01
→	Protège	0,32	0,04
→	Écoute	0,05	NS
Immaturité →	Comprenne	0,3	0,06
→	Soutienne	0,36	0,02
→	Console	0,32	0,04
→	Protège	0,3	0,06

CONCLUSION

On voit qu'il est tout à fait possible de faire des recherches évaluatives sur les psychothérapies à condition de respecter leur spécificité, et de prendre en compte les interrogations des cliniciens. Si la dimension heuristique de tels travaux est indéniable, il faut néanmoins rester modeste lors de la généralisation des résultats. À l'heure actuelle, ce type de recherche, du fait de son caractère non inférentiel, est davantage générateur de nouvelles hypothèses que de conclusions définitives.

BIBLIOGRAPHIE

AAPRO N., DAZORD A., GERIN P., DECOULON N., ANDREOLI A. — Institution, thérapies et formation des thérapeutes. *Jour. des Psyc.*, *114*, 28-30, 1994.

CARMINES E. C., ZELLER A. — Reliability and Validity Assessment. Quantitative Applications in the Social Sciences. *Beverly Hills/London Sage*, *5-71*, 1979.

DAZORD A., GERIN P., ANDREOLI A., REITH B., IAHNS J. F., ABENSUR J. — Pretreatment and Process Measures in Crisis Intervention as Predictors of Outcome. *Psychotherapy Research*, *1 (2)*, 1-13, 1991.

DAZORD A., GERIN P., REITH B., IAHNS J. F., ANDREOLI A. — Intervention de crise : processus de focalisation et évolution à long terme des patients. *L'Encéphale*, *XVIII (6)*, 639-645, 1992.

DAZORD A., MANIFICAT S., GERIN P., BEYER H., PECHINE C., GUILLAUD-BATAILLE J. M. — Les difficultés de l'enfant : représentations de la famille et des soignants; intérêt d'une enquête. *La Psychiatrie de l'Enfant*, *XXXVI*, 151-176, 1993.

DAZORD A., GERIN P. — Évaluation des psychothérapies : méthodologies et résultats. *L'information psychiat.*, *69 (5)*, 465-471, 1993.

GARFIELD S. L., BERGIN A. E. — *Handbook of Psychotherapy and Behavior Change*. J. Wiley and Sons, New York, 1986.

HORVATH A. O., SYMONDS B. D. — The Relation between Working Alliance and Outcome in Psychotherapy : a Metaanalysis. *J. Counsel Psychol.*, *38 (2)*, 139-149, 1991.

GERIN P., DAZORD A., SALI A. — *Psychothérapies et changements, méthodologie de leur évaluation*. PUF, collection NODULES, Paris, 1991.

GERIN P., DAZORD A. — Recherches cliniques «planifiées» sur les psychothérapies. Collection *Analyses et prospective*. INSERM/La Documentation Française, 1992a.

GERIN P., DAZORD A., GUISTI P., SALI A., MARIE-CARDINE M. — L'évaluation des facteurs psychothérapeutiques de changement en hôpital psychiatrique. *Psychothérapies*, *2*, 123-132, 1992b.

GERIN P., DAZORD A., BOISSEL J. P., CHIFFLET R. — Quality of Life Assessment in Therapeutic Trials : Rationale for and Presentation of a more Appropriate Instrument. *Fundam Clin. Pharmacol.*, *6*, 263-276, 1992c.

GLANTZ S. — *Primer of Biostatistics*. MacGraw-Hill Company Ed., 1981.

LAMBERT M. J., SHAPIRO D. A., BERGIN A. E. — The Effectiveness of Psychotherapy. In : GARFIELD S. L. & BERGIN A. E. (Eds), *Handbook of Psychotherapy and Behaviour Change*. New York, 1986.

LUBORSKY L., CRITS-CHRISTOPH P., MINTZ J., AUERBACH A. — *Who Will Benefit from Psychotherapy? Predicting Therapeutic Outcomes*. Basic Books, New York, 1988.

ORLINSKY D. — Apport de la recherche à la connaissance des psychothérapies. In : GERIN P.; DAZORD A. ed. *Recherches planifiées sur les psychothérapies*. Collection Analyses et prospective. INSERM/La Documentation Française, 1992.

SEULIN C., GERIN P., AMAR A., DUCLOS A., DAZORD A. — Présentation d'une échelle d'engagement du patient dans la relation soignant-soigné. *L'information psychiatrique*, *9*, 915-923, 1989.

STILES W., SHAPIRO D., ELLIOTT R. — Are all Psychotherapies Equivalent? *Am. Psychol.*, *41*, 165-180, 1986.

VOLLE M. — *Analyse des données*. Collection Économie et statistique avancées. Economica, Paris, 1985.

19

ÉVALUATION DES THÉRAPIES COMPORTEMENTALES EN PRATIQUE COURANTE

par V. Leclerc et R. Dardennes

L'évaluation est un principe intrinsèque aux thérapies comportementales. En effet, celles-ci reposent sur l'idée que les comportements pathologiques ou dysfonctionnels sont les troubles mêmes à modifier, et sont appris et maintenus par l'action de l'environnement. L'observation des comportements est donc essentielle à l'intervention thérapeutique et leurs modifications reflètent l'efficacité du traitement comportemental. La quantification des comportements est un élément important de ce type de traitement. Les thérapies comportementales sont des méthodes de type technologique; elles s'attachent à définir de la manière la plus opératoire possible les problèmes dont souffrent les patients et les techniques thérapeutiques employées à les résoudre s'efforcent d'objectiver leurs effets par la mesure répétée de comportements ou de l'anxiété éprouvée par le sujet et de démontrer la reproductibilité de leur efficacité en réalisant des essais thérapeutiques. Sur le plan pratique, l'évaluation présente deux aspects importants pour un thérapeute comportementaliste :

1) Que peut-on tirer des recherches cliniques sur l'efficacité des traitements comportementaux pour la pratique quotidienne?

2) Comment apprécier l'évolution de l'état du patient que l'on traite et savoir si l'amélioration observée est imputable au traitement effectué?

Un dernier aspect de l'évaluation des thérapies comportementales deviendra capital dans un futur proche : l'évaluation du rapport coût/bénéfice des traitements comportementaux et de l'adéquation de leurs modalités de délivrance par un rapport aux besoins de santé d'une population.

RETOMBÉES PRATIQUES DE LA RECHERCHE CLINIQUE

En 1981, un éditorial de Barlow (1981), éminent comportementaliste, était assez pessimiste à ce sujet : «La recherche clinique a une influence nulle ou faible sur la pratique clinique.» Certes, ce constat n'est pas propre à la thérapie comportementale mais il est particulièrement saillant dans une discipline qui a un parti pris de scientificité. Hormis les essais thérapeutiques de techniques comportementales, dont nous verrons plus loin certains obstacles quant à leur impact sur la pratique, deux types d'études ont des retombées potentielles sur la pratique usuelle.

D'abord, une thérapie comportementale est généralement la combinaison de plusieurs techniques : relaxation, exposition, auto-observation, pratique d'exercices à domicile (*homework*), etc. Si la plupart des études sur l'efficacité des thérapies comportementales évalue des «packages» thérapeutiques, il n'est pas inintéressant de rechercher quelles composantes sont particulièrement efficaces. Ces études, dites de démantèlement, permettent ainsi de simplifier une méthode en éliminant les éléments non contributifs ou au contraire de montrer la pertinence d'éléments supplémentaires. C'est la démarche entreprise, à titre d'exemple, par Turner, Ward et Turner (1979) pour déterminer les «ingrédients» thérapeutiques du traitement comportemental de la dépression conçu par Lewinsohn. Ces auteurs ont ainsi montré, lors de cette étude contrôlée, le surcroît d'efficacité obtenu par un programme visant à augmenter la participation du déprimé à des activités plaisantes par rapport au simple auto-enregistrement des activités effectuées par le sujet. À l'inverse, nous citerons le travail d'Edelman et Chambless (1993) sur l'effet propre de la pratique, entre les séances de psychothérapie, d'auto-expositions chez les agoraphobes. En effet, il est généralement admis que la pratique d'exercices à domicile est une composante thérapeutique importante. Cependant, peu d'auteurs ont cherché à vérifier cette affirmation. Dans cette étude contrôlée, mais non randomisée, la pratique d'exercices à domicile n'apporte aucun avantage thérapeutique à court terme. Le débat reste ouvert, vu le faible nombre de travaux dans ce domaine et leurs résultats discordants, mais peut avoir des conséquences pratiques dans le sens d'une simplification de la prise en charge des agoraphobes.

Un deuxième point porte sur la généralisation des résultats publiés. L'évaluation des thérapies comportementales est souvent réalisée par des équipes chevronnées au sein d'unités spécialisées, sur une population sélectionnée de patients. Peut-on extrapoler les résultats obtenus dans ces conditions à l'application de ces méthodes en situation ordinaire? Il n'est plus alors question d'homogénéiser au maximum les paramètres susceptibles de gêner l'appréciation de l'effet propre au traitement, mais plutôt d'évaluer sa robustesse, sa capacité à être «tout-terrain». Le compte-rendu des résultats obtenus en pratique courante souffre bien évidemment de faiblesses méthodologiques; l'efficacité observée ne peut être strictement imputée au traitement et les biais de publication favorisent la communication préférentielle des bons résultats. Néanmoins, de tels exposés permettent d'apprécier la faisabilité et l'efficacité de la prise en charge de sujets moins sélectionnés. Ainsi, les séries de patients obsessionnels rapportées par Kirk (1983) et Buisson *et al.* (1989) permettent d'estimer qu'un sujet sur 10 abandonne le traitement et 6 à 8 sont nettement améliorés; les résultats les plus faibles sont observés par l'équipe recourant à une évaluation indépendante du thérapeute (Buisson *et al.*, 1989). De la même manière, Samuel-Lajeunesse *et al.* (1985) montrent l'efficacité en routine des techniques assertives de groupe sur une série de 31 patients hétérogènes avec une nette amélioration dans deux tiers des cas. Ces quelques exemples permettent d'apprécier la pertinence des thérapies comportementales en pratique courante avec des patients (et des thérapeutes) aux caractéristiques variées.

Mais seules les études contrôlées, malgré leur difficulté, permettent de démontrer l'efficacité des techniques utilisées. La question alors posée par le praticien est la pertinence des résultats publiés quant à son exercice, c'est-à-dire la notion de résultat est-elle la même pour un chercheur et un praticien?

Tout d'abord, il est important de savoir si les (bons) résultats publiés concernent la population de départ ou uniquement les sujets qui ont mené le traitement à son terme.

Le fait de ne pas prendre en compte les sujets ayant abandonné le traitement donne certes des résultats plus flatteurs mais expose le praticien à la déception de découvrir une efficacité globalement inférieure à celle qu'il a lue. Une analyse des études sur le traitement des troubles panique avec agoraphobie effectuée par Michelson et Marchione (1991) montre ainsi que 75 % des sujets menant leur traitement à terme y répondent, mais qu'un sur cinq l'abandonne; le véritable taux de réponse n'est plus alors que de 60 %, estimation plus modeste mais plus réaliste.

Un deuxième problème concerne la diversité des outils d'évaluation de l'état du sujet traité. Or ces outils sont loin d'être équivalents. Ogles *et al.* (1990) ont réalisé une méta-analyse sur l'influence du type d'instrument d'évaluation sur le résultat thérapeutique observé dans l'agoraphobie. Une revue de 106 études a permis de recenser 98 instruments différents! De plus l'importance du résultat thérapeutique varie selon l'outil utilisé dans un rapport de 1 à 6, même s'il s'agit d'instruments appartenant à la même classe (ex. questionnaires d'auto-évaluation). L'estimation de l'ampleur du changement obtenu est donc fonction de la manière dont on l'évalue; comme il est difficile de connaître l'ensemble des moyens d'évaluation pour un problème donné et que les études comparatives sont rares, on ne peut être certain de retrouver dans la pratique la même intensité de changement que celle publiée si les indices d'efficacité utilisés dans chaque cas ne sont pas les mêmes.

Or, l'importance du changement obtenu par le traitement est bien plus significatif pour la pratique que le degré de signification statistique de la comparaison des moyennes de deux groupes; plus que le changement moyen, c'est le changement obtenu pour un individu qui est critique. Jacobson et Truax (1991) ont proposé une manière statistique d'apprécier l'effet individuel d'une psychothérapie dans les essais thérapeutiques, l'index de changement fidèle (*Reliable Change Index*) qui permet d'apprécier si la variation de score observée chez un patient donné est significative, c'est-à-dire supérieure aux erreurs de fluctuation liées à la mesure. Un tel index, calculable pour des instruments d'évaluation standardisés tels les autoquestionnaires (Beidel *et al.*, 1993), peut alors permettre une appréciation individuelle plus précise que la validation du changement obtenu par le jugement global du patient, de ses proches ou encore du thérapeute; cet index est plus sensible que de juger la réponse au traitement en vérifiant si les scores qui en résultent ne sont plus dans les normes «pathologiques» et sont alors dans les limites de la normale.

Enfin, l'évolution même de la recherche clinique peut avoir des conséquences, inattendues, sur la pratique clinique. On peut en effet noter une médicalisation de l'abord comportemental, un passage de la modification d'un comportement donné par un traitement *ad hoc* après analyse à un modèle médical où il s'agit de traiter une catégorie de sujets — c'est-à-dire une affection — par un traitement standardisé. Ce glissement remet implicitement en question l'utilité fondamentale de l'analyse fonctionnelle du comportement, les sujets atteints d'une même affection ayant suffisamment de caractéristiques communes pour prédire leurs réponses à un traitement non personnalisé. Ce point a été souligné récemment par Emmelkamp (1992) qui cite à l'appui de cette hypothèse des travaux où le fait d'échanger les analyses fonctionnelles de sujets différents mais souffrant d'un même trouble n'a pas d'incidence significative sur le résultat thérapeutique : le fait de traiter M. Dupont sur la base de l'analyse fonctionnelle effectuée chez M. Durand serait sans conséquence sur le résultat obtenu! Ces données, très provocatrices, méritent d'être vérifiées par d'autres équipes et soulignent à nouveau les retombées pratiques d'une recherche clinique quantifiée.

ÉVALUATION : COMPOSANTE DE LA MÉTHODE THÉRAPEUTIQUE

Comment apprécier l'évolution de l'état du patient que l'on traite et savoir si l'amélioration observée est imputable au traitement effectué? Une approche quantitative du comportement permet de mieux définir la cible du traitement, juger son importance (fréquence, durée, intensité) et sa stabilité (répétition de la mesure). Il s'agit donc de définir un indice sur lequel on appréciera l'évolution de l'état du sujet. Plusieurs sources information sont disponibles pour définir puis mesurer cet indice : le récit du patient, l'observation directe par le thérapeute au fil de la consultation (ex. : fréquence d'un bégaiement) ou lors de mises en situation (jeux de rôle, tests d'approche comportementale, tests *in vivo*), l'observation par des tiers (ex. : pour les enfants) ou dans certains cas des mesures physiologiques.

Deux méthodes permettent de quantifier facilement l'état du patient : les questionnaires et l'auto-enregistrement des données par le sujet lui-même. Les questionnaires donnent le plus souvent une mesure globale de l'état du patient et sont de ce fait peu sensibles à une intervention ciblée sur un comportement donné. Plusieurs études ont ainsi montré cette faible corrélation entre différents types d'évaluation, par exemple, l'étude de Burkhart *et al.* (1979) où les corrélations entre les questionnaires d'assertivité et l'évaluation d'un jeu de rôle-test sont faibles et sont nulles entre les mesures précédentes et une évaluation en situation réelle (un comparse demandait au sujet des services de plus en plus excessifs). Les questionnaires sont également sujets aux biais de mémorisation qui affectent les réponses lorsque les passations sont répétées. Leur utilisation se limite donc à des évaluations ponctuelles, avant et après traitement ainsi que dans le suivi à long terme pour vérifier la stabilité des résultats obtenus.

Par contre, l'auto-enregistrement, permet, dans certaines limites, la réalisation de mesures multiples. Dans ce cas, il est demandé au patient d'identifier et d'enregistrer lui-même la survenue d'un comportement, d'une de ses caractéristiques ou de ses conséquences. Cette méthode est peu coûteuse, permet une observation plus complète — quotidienne et quasi continue —, et demeure parfois la seule source de données pour certains problèmes (troubles sexuels ou du sommeil, comportements couverts comme les ruminations du déprimé, les pensées obsédantes, ou encore les inquiétudes d'un anxieux généralisé). Pour ce faire, l'auto-enregistrement doit être adapté aux caractéristiques du comportement-cible afin de faciliter le recueil.

– Le plus souvent, l'auto-enregistrement est de type narratif : le patient tient un agenda où il note la survenue d'un comportement donné et décrit ses caractéristiques (ex. : intensité, fréquence, durée) et son contexte. Ce carnet peut être structuré en rubriques (heure, contexte, intensité, durée, etc.) mais il est souvent utile alors de laisser une place pour les remarques additionnelles du patient.

– Si le comportement est bien défini dans le temps et a une durée stable, il est possible d'évaluer seulement sa fréquence quotidienne : nombre de cigarettes fumées, nombre de lavages de mains pour un obsessionnel ou d'appels téléphoniques à ses enfants pour une anxieuse généralisée préoccupée par leur santé.

– La durée du comportement peut être l'indice principal à enregistrer si celle-ci varie ou est la cible du traitement : on peut ainsi faire mesurer le temps pris par la toilette ou les vérifications, le temps de travail ou le nombre de pages écrites par un thésard en difficulté ou encore la durée des sorties d'un agoraphobe.

– Une difficulté peut être d'évaluer des comportements trop fréquents pour être comptabilisés de manière exhaustive : tics, onychophagie ou trichotillomanie, etc.

On peut alors procéder à un échantillonnage en limitant l'auto-enregistrement à une tranche horaire prédéfinie ou à des recueils ponctuels (*spot-checking*) en notant à différentes heures de la journée si le comportement est présent ou pas.

– L'auto-enregistrement peut également consister en évaluations répétées de l'humeur du sujet ou de l'intensité de l'anxiété qu'il ressent, sur une échelle de 0 à 10 par exemple, à heure fixe ou de manière ponctuelle à plusieurs temps différents de la journée.

Enfin, l'évaluation peut s'appuyer non sur le comportement lui-même mais sur ses conséquences ou son résultat. Les exemples les plus évidents sont la longueur des ongles chez un onychophage ou le poids chez un hyperphage.

L'auto-enregistrement soulève deux questions : ce type d'évaluation est-il précis? Le fait d'inciter un sujet à s'observer modifie-t-il son comportement? La précision absolue de l'auto-enregistrement est difficile à établir; d'une part, il est difficile de corréler des comportements couverts à des comportements ouverts. D'autre part, les divergences observées entre une auto-observation et une observation externe peuvent s'interpréter différemment, l'une des deux s'avérant meilleure que l'autre selon la nature du comportement observé (Nelson, 1977). Par contre, certains facteurs influençant cette précision sont connus. Le fait de vérifier explicitement la précision de l'auto-enregistrement effectué par un sujet augmente la qualité de la mesure. Il en est de même pour le renforcement positif de l'exactitude de la mesure. Il est plus facile de monitorer des comportements moteurs que des comportements verbaux (comme de noter la fréquence des *« vous savez ce que c'est »*). L'enregistrement sera d'autant plus précis qu'il se fera sans délai. En pratique, il est important de fournir au patient des instructions claires sur ce qui lui est demandé à l'aide d'exemples, d'en vérifier leur compréhension, tout en sachant qu'il faudra un minimum d'entraînement au sujet pour parvenir à un résultat correct. Le fait de comparer ouvertement les données de l'auto-enregistrement à une mesure (ex. : le poids pour un relevé des prises alimentaires) améliorera la qualité de leur recueil.

Mais si l'on peut améliorer la précision de l'auto-enregistrement, celui-ci en tant que tel modifie également la fréquence du comportement qu'il doit mesurer. Cette réactivité de la mesure s'apparente souvent à un effet thérapeutique car son effet dépend de la désirabilité des comportements évalués : le monitorage augmente les comportements désirés et diminue les comportements non désirés. Cette propriété est plus gênante dans le cadre de la recherche clinique que dans la pratique; son effet se confond avec celui propre à l'intervention thérapeutique et peut être recherché pour accroître le changement. Plusieurs facteurs affectent cette réactivité (Nelson, 1977, 1981); le fait de fixer des objectifs comportementaux à atteindre l'augmente, de même que le renforcement positif de la réalisation de ces objectifs. Le moment de l'enregistrement peut avoir son importance; enregistrer avant de fumer ou de grignoter plutôt qu'après entraîne une plus grande diminution de ces comportements. La réactivité est meilleure si l'auto-enregistrement est continu et ne concerne qu'un seul comportement.

L'auto-enregistrement est donc un outil particulièrement pratique; au prix de quelques précautions, il permet une quantification répétée et suffisante pour apprécier l'évolution du comportement; il peut également jouer un rôle actif dans la modification du comportement.

L'évaluation du comportement permet d'observer son évolution sous traitement mais peut-on pour autant imputer le changement obtenu chez un patient donné au traitement qui lui a été appliqué?

Une approche expérimentale consiste à mettre en place un protocole à cas unique. Une observation intensive et répétée du patient est effectuée avant, pendant et après le traitement. Le protocole consiste à introduire puis à retirer une ou plusieurs techniques thérapeutiques et à comparer l'évolution avec et sans traitement donné. Ces protocoles, dits ABA, ABAB, ABCBC, ..., n'ont d'intérêt que si le comportement observé est très malléable et si le traitement n'a qu'un effet temporaire; ils donnent de bons arguments en faveur de la causalité du traitement mais leur validité est uniquement restreinte au cas étudié. En dépit de leur relative simplicité, ces protocoles sont très peu utilisés en pratique courante. De fait, l'imputabilité ne peut le plus souvent être démontrée, elle est plausible. Cette plausibilité sera d'autant plus forte que les autres hypothèses pouvant expliquer le changement observé seront moins plausibles. Plusieurs éléments permettent d'étayer l'effet probable du traitement. Des évaluations répétées avant traitement (ligne de base) permettent d'apprécier la stabilité du comportement, sa plage de variation, l'éventuel effet de l'auto-observation et une régression à la moyenne des mesures. L'anamnèse du sujet et les données épidémiologiques donnent des indications supplémentaires sur l'évolution spontanée du phénomène, par exemple l'absence de rémission spontanée. Un délai d'action court et un changement important sont des arguments supplémentaires en faveur d'un effet probable du traitement, ainsi que l'efficacité de la technique utilisée sur des sujets hétérogènes (homme ou femme, adulte ou sujet âgé, universitaire ou faiblement scolarisé, etc.).

Kazdin (1981) définit ainsi 3 types d'études de cas par ordre croissant d'imputabilité du changement observé au traitement :

1) Évaluation avant-après traitement.

2) Évaluation répétée.

3) Évaluation répétée et cas multiples hétérogènes.

La pratique de mesures répétées du comportement est, en définitive, le meilleur garant pour s'assurer en routine de l'effet d'un traitement comportemental.

VERS UNE ÉVALUATION DES MODALITÉS DE DÉLIVRANCE DE LA THÉRAPIE COMPORTEMENTALE

L'évaluation de la qualité des soins est devenue ces dernières années un sujet crucial étant donné l'accroissement important des dépenses de santé. Or les thérapies comportementales ont connu un développement très important ces dernières années. Ainsi, à la Clinique des maladies mentales et de l'encéphale, le nombre d'actes de thérapie comportementale est passé en quinze ans de 400/an à plus de 3000/an. Il est donc probable qu'à terme, les psychothérapies comportementales telles qu'elles sont réellement pratiquées seront soumises à une évaluation de leur rapport coût-efficacité et de l'adéquation de leurs conditions de délivrance aux besoins de la population. Dans cette perspective, Parry (1992) passe en revue les différentes méthodes d'évaluation des services de psychothérapie et propose un cadre de réflexion basé sur les 6 critères de Maxwell : pertinence des services offerts par rapport aux besoins de la population, équité dans l'accès aux soins, accessibilité et acceptabilité du traitement, efficacité et enfin productivité. La prochaine étape de l'évaluation en thérapie comportementale sera de développer des outils permettant de répondre à ces

objectifs. Ces exigences rendront plus que jamais nécessaire une recherche clinique soucieuse de généraliser ces résultats à la pratique quotidienne.

Bibliographie

AGATHON M., SAMUEL-LAJEUNESSE B. — Bilan de quatre années de thérapie comportementale. *Psychol. Méd., 13*, 107-119, 1981.

ALLISON D. B., GORMAN B. S. — Calculating Effect Sizes for Meta-Analysis : the Case of the Single Case. *Behaviour Research and Therapy, 31*, 621-631, 1993.

BARLOW D. H. — On the Relation of Clinical Research to Clinical Practice : Current Issues, New Directions. *Journal of Consulting and Clinical Psychology, 49*, 147-155, 1981.

BEIDEL D. C., TURNER S. M., COOLEY M. R. — Assessing Reliable and Clinically Significant Change in Social Phobia : Validity of the Social Phobia and Anxiety Inventory. *Behaviour Research and Therapy, 31*, 331-337, 1993.

BUISSON G., PATILLOT P., LECLERC V., AGATHON M., SAMUEL-LAJEUNESSE B. — Thérapie comportementale et symptomatologie obsessionnelle. *Ann Psychiatr, 4*, 236-241, 1989.

BURKHART B. R., GREEN S. B., HARRISON W. H. — Measurement of Assertive Behavior : Construct and Predictive Validity of Self-Report, Role-Playing and *in vivo* Measures. *Journal of Clinical Psychology, 35*, 376-383, 1979.

COTTRAUX J., BOUVARD M., LÉGERON P. — *Méthodes et échelles d'évaluation des comportements*. EAP, Issy-les-Moulineaux, 1985.

EDELMAN R. E., CHAMBLESS D. L. — Compliance during Sessions and Homework in Exposure-Based Treatment of Agoraphobia. *Behaviour Research and Therapy, 31*, 767-773, 1993.

EMMELKAMP P. M. G. — Behaviour Therapy in the Fin de Siècle. In : COTTRAUX J., LÉGERON P., MOLLARD E. (eds). *Annual Series of European Research in Behavior Therapy*, vol. 6, Swets & Zeitlinger, Amsterdam, 149-163, 1992.

GUELFI J. D. (ed) — *L'évaluation clinique standardisée en psychiatrie*. Tome I. Éditions Médicales Pierre Fabre, Boulogne, 1993.

JACOBSON N. S., TRUAX P. — Clinical Significance : a Statistical Approach to Defining Meaningful Change in Psychotherapy Research. *Journal of Consulting and Clinical Psychology, 59*, 12-19, 1991.

KAZDIN A. E. — Drawing Valid Inferences from Case Studies. *Journal of Consulting and Clinical Psychology, 49*, 183-192, 1981.

KENDALL P. C., LIPMAN A. J. — Psychological and Pharmacological Therapy : Methods and Modes for Comparative Outcome Research. *Journal of Consulting and Clinical Psychology, 59*, 78-87, 1991.

KIRK J. W. — Behavioural Treatment of Obsessional-Compulsive Patients in Routine Clinical Practice. *Behaviour Research and Therapy, 21*, 57-62, 1983.

MICHELSON L. K., MARCHIONE K. — Behavioral, Cognitive and Pharmacological Treatments of Panic Disorder with Agoraphobia : Critique and Synthesis. *Journal of Consulting and Clinical Psychology, 59*, 100-114, 1991.

MORLEY S., ADAMS M. — Some Simple Statistical for Exploring Single-Case Time-Series Data. *British Journal of Clinical Psychology, 28*, 1-18, 1989.

MORLEY S., ADAMS M. — Graphical Analysis of Single-Case Time-Series Data. *British Journal of Clinical Psychology, 30*, 97-115, 1991.

NELSON R. O. — Assessment and Therapic Functions of Self-Monitoring. In : HERSEN M., EISLER R. M. & MILLER P. M. (Eds), *Progress in Behavior Modification*, vol. 5, Academic Press, New York, 263-308, 1977.

NELSON R. O. — Realistic Dependent Measures for Clinical Use. *Journal of Consulting and Clinical Psychology, 49*, 168-182, 1981.

NEZU A. M., PERRI M. G. — Social-Problem Therapy for Unipolar Depression : an Initial Dismantling Investigation. *Journal of Consulting and Clinical Psychology, 57*, 408-413, 1989.

OGLES B. M., LAMBERT M. J., WEIGHT D. J., PAYNE I. R. — Agoraphobia Outcome Measurement : a Review and Meta-Analysis. *Psychological Assessment, 2*, 317-325, 1990.

PARRY G. — Improving Psychotherapy Services : Applications of Research, Audit and Evaluation. *British Journal of Clinical Psychology, 31*, 3-19, 1992.

PARRY G., SHAPIRO D. A., FIRTH J. — The Case of the Anxious Executive : a Study from the Research Clinic. *British Journal of Medical Psychology, 59*, 221-233, 1986.

TURNER R. W., WARD M. F., TURNER J. — Behavioral Treatment for Depression : an Evaluation of Therapeutic Components. *Journal of Clinical Psychology, 35*, 166-175, 1979.

SAMUEL-LAJEUNESSE B., DEGLERIS N., BROURI R., AGATHON M. — Appréciation des effets des techniques assertives en groupe en milieu psychiatrique. *Ann. Médico. Psychol., 143*, 458-464, 1985.

20

LES INVENTAIRES COMPORTEMENTAUX CHEZ L'ENFANT EN POPULATION CLINIQUE ET EN POPULATION GÉNÉRALE

par M. P. Bouvard

La psychopathologie quantitative n'est pas une discipline récente chez l'enfant. On peut la faire débuter avec les travaux de Binet et Simon en 1905 qui, les premiers, tentèrent de quantifier les performances cognitives des enfants et furent à l'origine de la mesure du quotient intellectuel. Si les psychologues furent les précurseurs de la mesure chez l'enfant, il faudra attendre l'expansion chez l'adulte des instruments cliniques plus spécifiquement destinés à mesurer des syndromes cliniques ou des entités symptomatiques pour voir se développer chez l'enfant et l'adolescent de pareils instruments.

Le développement de la recherche a imposé, chez l'adulte d'abord, et plus récemment, chez l'enfant, la nécessité pour le clinicien de développer des instruments de mesure fiables, reproductibles et interprétables. La publication en 1985 d'un numéro spécial de *Psychopharmacology Bulletin* a montré que si l'emploi d'instruments de mesure quantitatifs de la symptomatologie ou du comportement n'était pas courant en France, ce n'était pas dû au manque d'instruments actuellement disponibles dans différents domaines de la pathologie mentale de l'enfant. Plus récemment, l'ouvrage de Lelord et Barthélémy (les échelles d'évaluation clinique en psychiatrie de l'enfant, 1991) confirmait le nombre important d'outils disponibles aujourd'hui.

Ces instruments sont soit directement issus d'adaptation d'échelles validées chez l'adulte, soit des traductions d'instrument mis au point en Angleterre et aux États-Unis, pour lesquels se pose la question de la validité culturelle de la traduction notamment en ce qui concerne les seuils pathologiques des comportements.

On peut distinguer différents niveaux de mesure en psychopathologie quantitative :

1) *Le niveau comportemental* : ce niveau correspond à ce qui est observable par toute personne, des attitudes, comportements, conduites du sujet. Il est censé être le plus objectif. Il discrimine entre des comportements dits externalisés à type d'agitation, d'agressivité ou d'opposition et des comportements dits internalisés qui correspondent à des attitudes de repli, d'inhibition. À ce titre, il n'est pas superposable aux classifications syndromiques, et a, dans ce domaine, un manque de spécificité. Ainsi, à un même type de comportements peuvent correspondre des entités syndromiques différentes. C'est ce niveau de mesure comportementale qui a

été le plus exploré chez l'enfant car l'axe comportement internalisé et comportement externalisé est celui qui est le plus consensuel actuellement.

2) *Le niveau syndromique* : ce niveau correspond à l'évaluation d'entités pathologiques reconnues telles que la dépression ou la schizophrénie. Il suppose un consensus sur la reconnaissance et l'identification de ces entités. Si ce consensus est clair concernant les troubles comme la dépression, certains troubles anxieux (phobies, anxiété de séparation...), il est moins évident pour d'autres groupes tels que l'hyperactivité, les troubles anxieux généralisés, l'hyperanxiété ou l'autisme infantile. Ainsi chez l'enfant, de nombreux instruments existent dans l'évaluation des syndromes, mais ils ne font pas encore l'objet d'une utilisation reconnue par tous.

3) *Le niveau étiologique* : ce niveau correspond à l'évaluation d'une cause ou de facteurs causaux directement impliqués dans la survenue d'un trouble psychiatrique. Il reste dans le domaine des troubles mentaux un niveau hypothétique et expérimental.

Nous verrons dans la suite de ce chapitre les instruments qui ont été développés dans l'évaluation comportementale des troubles de l'enfant.

ASPECTS MÉTHODOLOGIQUES ET MÉTROLOGIQUES

Les inventaires comportementaux chez l'enfant correspondent à l'évaluation de comportements observables, durables et répétitifs de l'enfant. Néanmoins le terme de comportemental isolé paraît insuffisant car s'il tente d'objectiver des attitudes de l'enfant, il fait référence à des modèles parallèles de fonctionnement psychologique et il apparaît plus approprié de parler d'inventaires psychocomportementaux. Ces inventaires sont basés sur l'observation de l'enfant en milieu naturel et font appel à des sources issues de l'environnement habituel de l'enfant. C'est ainsi qu'habituellement on distingue des formes destinées à l'évaluation en milieu familial, des formes destinées à l'évaluation en milieu scolaire et pour certains de ces instruments, des formes qui ont été développées afin de permettre une évaluation comportementale dans d'autres environnements.

Ces inventaires comportementaux peuvent être séparés en trois types différents en fonction de la nature de la mesure :

1) ceux qui évaluent directement la sémiologie comportementale, qui tentent de différencier, de définir et de quantifier différents comportements observables de l'enfant ;

2) ceux qui évaluent de façon précise les situations dans lesquelles surviennent les perturbations comportementales ; ces derniers s'intéressent particulièrement au cadre dans lequel apparaissent les attitudes pathologiques et restent très globaux concernant le type de comportement pathologique à proprement parler ;

3) enfin ceux qui s'intéressent au retentissement d'un trouble en terme de qualité d'adaptation dans les domaines scolaires ou relationnels et axent leur mesure sur la notion de retentissement plus que sur l'aspect vraiment sémiologique.

L'interprétation des résultats fait référence à deux types de scores obtenus à partir de ces inventaires ;

1) des scores globaux qui correspondent à un indice de psychopathologie générale, non spécifique, et qui expriment l'intensité du trouble ;

2) des scores factoriels issus des analyses statistiques des instruments durant leur étape de validation qui sont des indices plus précis concernant différents sous-groupes de comportement pathologique.

Il reste qu'il existe de nombreux biais dans l'évaluation comportementale qui doivent être pris en compte dans leur interprétation. Ainsi, il existe une surestimation des comportements extravertis de type agitation, troubles des conduites, par rapport aux comportements introvertis de type repli, inhibition. De cette façon les premiers sont plus facilement repérés, plus facilement cotés que les seconds.

Barkley (1980) donnait les principales qualités métrologiques que devaient avoir ces inventaires.

– la clarté et spécificité des énoncés de chacun des items afin d'éviter le phénomène de redondance ;

– l'existence d'un nombre suffisant d'items. Ce nombre doit être assez important pour permettre une validité comportementale ou psychologique de la mesure mais pas trop important afin d'éviter des délais de remplissage prolongés qui introduisent là aussi des biais de cotation ;

– une quantification précise et suffisante de chacun des items de ces inventaires. Une réponse purement qualitative en oui/non apparaît peu discriminante pour distinguer les populations cliniques et non cliniques, et il est préférable en fonction du but, d'y substituer une qualification avec des degrés de sévérité plus nombreux (de 0 à 4, de 0 à 7) associés à une description précise de chacun de ces degrés ;

– les caractères métrologiques des inventaires : les différentes caractéristiques sont communes à tous les instruments d'évaluation :

- la validité externe de l'instrument par rapport à d'autres instruments de mesure ;
- la validité discriminante de l'instrument ;
- la validité prédictive est particulièrement importante chez l'enfant en terme de développement ;
- la validité inter-juges, les tests-retests ;

– les conditions d'emploi et l'intérêt général de l'inventaire doivent être précisés en distinguant ceux qui ont une validité en terme de profil comportemental et ceux qui sont destinés particulièrement à suivre l'évolution de l'enfant et l'efficacité éventuelle des thérapeutiques engagées.

PRINCIPAUX INVENTAIRES COMPORTEMENTAUX CHEZ L'ENFANT

Nous n'aborderons pas dans le détail la liste exhaustive des inventaires comportementaux disponibles chez l'enfant et l'adolescent. On pourra se référer à l'ouvrage publié par Barthélémy et Lelord, 1991, ou à l'article publié dans l'encyclopédie médico-chirurgicale (Dugas, Bouvard, 1994) pour avoir une bibliographie plus exhaustive. Après avoir évoqué les principaux instruments dans différents domaines de comportement, nous aborderons en détail deux de ces instruments — les échelles de Conners et l'inventaire de comportement d'Achenbach — en illustrant leur utilisation d'exemples cliniques.

Les différents types d'instruments

Les inventaires comportementaux.

Les principaux inventaires comportementaux utilisés actuellement sont :

– la *Child Behavior Checklist* (Achenbach, Edelbrock,1983) ;

– les échelles de Conners (Conners, 1982) : il en existe deux formes : une forme parent et une forme enseignant ;

– le questionnaire d'évaluation comportementale de Rutter (1971) : il en existe également deux formes : une forme parent et une forme enfant ;

– les questionnaires d'évaluation de Devereux pour enfants et adolescents.

Les questionnaires de situation

Les questionnaires les plus connus sont ceux de Barkley (1980) dont il existe trois formes : pour la maison, pour l'école et pour les institutions médicalisées.

Les échelles d'adaptation

La plus utilisée est la *Children Global Assessment Scale* (Shaffer, 1983). Cette échelle est une adaptation de la *Global Assessment Scale* validée chez l'adulte et dont une forme a été réalisée pour être utilisable en tant que axe V du DSM-III-R (1987) (la *GAF Global Assessment Functioning*).

Les échelles symptomatiques globales

La forme la plus utilisée est la *Children Psychiatric Rating Scale* (Fish, 1968) qui est une échelle à 63 items basée sur l'observation directe de l'enfant et qui distingue 15 facteurs différents de la psychopathologie de l'enfant (facteurs psychotiques, l'hostilité, l'hyperactivité, l'anxiété, les troubles de la pensée, les symptômes névrotiques, la dépression, l'excitation, le retrait, les conduites antisociales, les symptômes organiques, les troubles de la parole, les troubles du sommeil, les troubles de l'alimentation, l'énurésie et l'encoprésie).

La CBCL et les questionnaires de Conners

Ces instruments sont présentés en annexe.

La Children Behavior Checklist (Achenbach, Edelbrock, 1978, 1979, 1983) (CBCL)

• *Présentation générale de l'instrument*

La CBCL est actuellement l'inventaire comportemental le plus utilisé dans les études épidémiologiques en population non clinique. Il s'agit d'un instrument qui évalue les comportements d'enfants et d'adolescents de 4 à 16 ans, mais deux formes ont été développées, une forme pour le jeune enfant évaluant les comportements précoces de 1 à 3 ans, et une forme pour le grand adolescent de 16 à 19 ans. Ce questionnaire est rempli par les parents même si une forme « enseignant » est actuellement en développement.

Il comporte 138 items répartis en deux grandes catégories. La première catégorie comprend 20 items qui évaluent les compétences de l'enfant dans différentes activités : vie quotidienne, relations sociales et performances scolaires. La deuxième catégorie comprend 118 items qui correspondent à des comportements différents regroupés en facteurs. Chaque item est coté de 0 à 2 en fonction de sa présence et de fréquence. L'analyse de la CBCL peut se faire de trois façons :

1) par la mesure d'un score global qui correspond à un index de gravité psychopathologique ;

2) par l'analyse des scores factoriels issus d'une analyse en composantes principales. 9 facteurs ont été isolés : a) le facteur psychotique, b) le facteur anxiété-dépression, c) le facteur trouble de la communication, d) le facteur obsessionnel compulsif, e) le facteur plaintes somatiques, f) le facteur retrait social, g) le facteur hyperactivité, h) le facteur comportement agressif, i) le facteur comportement délinquant ;

3) le 3e type d'analyse correspond à deux scores : le score internalisé, correspondant aux quatre premiers facteurs, et le score externalisé qui correspond aux trois derniers facteurs.

La CBCL a fait l'objet d'analyses statistiques et de validation pour chacun de ses scores factoriels, intermédiaire ou global, qui permettent d'aboutir à des données normatives précises. Cet inventaire a été traduit en français et a fait l'objet d'études de validation (Fombonne, 1988, 1989). Il est surtout utilisé et valide en population générale. Il existe actuellement des études en cours sur sa validité en milieu clinique.

En population générale, la CBCL a une sensibilité d'environ 70 %, une spécificité pour chacun des facteurs d'environ 75 %. Elle est particulièrement intéressante dans les études épidémiologiques.

• *Exemple d'utilisation de la CBCL : l'étude de Chartres (Fombonne, 1994)*

Dans cette étude, les auteurs ont évalué 2 441 enfants d'un département en deux étapes :

– la première étape consistait à faire remplir aux parents et enseignants d'enfants scolarisés d'une part la CBCL, d'autre part le questionnaire de Rutter (1971) ;

– la deuxième étape a consisté à comparer les enfants qui avaient des scores élevés sur différents facteurs de la CBCL avec une évaluation clinique basée sur l'utilisation du DSM-III.

La première étape a montré une corrélation élevée entre les scores obtenus dans les questionnaires parents et ceux des questionnaires enseignants. Cette corrélation est particulièrement élevée en terme quantitatif. Il existe par contre un taux de discordance plus élevé si l'on considère le repérage des différents domaines de comportement en fonction de la source. Ainsi, les comportements externalisés sont plus repérés en milieu scolaire, alors que les comportements internalisés sont plus repérés en milieu familial.

Les résultats de la deuxième étape sont présentés dans le tableau 1.

Tableau 1. *Concordance entre l'évaluation clinique et la CBCL*

CBCL	*Évaluation clinique en %*			
	Troubles émotionnels	*Troubles des conduites*	*Tout type de trouble*	*Absence de trouble clinique*
Parent seul	35,3	4,5	17,9	29,8
Enseignant seul	4,9	18,2	12,8	18
P + E	41,2	77,3	61,5	11,8
P ou E	82,4	100	92,3	59,5

Ces données confirment qu'il existe un très bon taux de concordance entre l'évaluation clinique et comportementale des troubles externalisés, puisqu'il existe une concordance de 100 % avec l'évaluation des parents ou des enseignants. Cette concordance est moins bonne concernant les troubles émotionnels puisque 20 % des sujets présentant des troubles émotionnels ne sont reconnus ni dans le questionnaire parent ni dans le questionnaire enseignant. Pour le questionnaire parent seul, elle est plus forte pour les troubles émotionnels que pour les troubles des conduites et à l'inverse, pour le questionnaire enseignant seul, plus forte pour les troubles des conduites que pour les troubles émotionnels.

Le taux de concordance est plus faible en ce qui concerne les populations d'enfants ne présentant pas de diagnostic clinique d'un trouble psychiatrique puisqu'il est estimé à 60 % dans cette étude, ce qui signifie que plus d'un tiers des enfants ne présentant pas de pathologies mentales reconnaissables à travers les catégories du DSM-III ont des scores élevés dans un ou plusieurs facteurs de la CBCL.

Cet instrument confirme, à travers cette étude, ses bonnes qualités pour l'évaluation et le dépistage de perturbations comportementales dans des populations non cliniques telles la population générale ou les populations pédiatriques.

Les échelles de Conners (1969, 1970, 1982)

• *Présentation générale des questionnaires de Conners*

Les échelles de Conners ont été développées en plusieurs étapes par Conners (1969, 1970, 1982) et Goyette *et al.* (1978). Elles sont destinées à évaluer principalement la symptomatologie de l'hyperactivité et des troubles comportementaux qui lui sont associés. Ces questionnaires sont basés sur l'observation directe de l'enfant par l'entourage, chaque item étant coté de 0 à 3 selon l'intensité de ces items. Il existe 3 versions des questionnaires de Conners :

– Le questionnaire destiné aux parents (*Parents Symptoms Questionnaire*, PSQ). La forme révisée de ce questionnaire comprend 48 items dont les scores se répartissent en 5 facteurs :

1) troubles des conduites,
2) problèmes d'apprentissage,
3) manifestations psychosomatiques,
4) impulsivité-hyperactivité,
5) anxiété.

En outre, il existe un index d'hyperactivité correspondant aux items les plus caractéristiques de l'hyperactivité de l'enfant de 10 items.

– Le questionnaire destiné aux enseignants (*Teacher Symptoms Questionnaire*, TSQ). Ce questionnaire comprend 28 items et est rempli par les enseignants. Sa structure factorielle comprend trois facteurs :

1) troubles des conduites,
2) hyperactivité,
3) immaturité/passivité.

Comme dans le questionnaire destiné aux parents, il existe un index d'hyperactivité constitué de 10 items.

– Le troisième questionnaire est l'individualisation de l'index d'hyperactivité de 10 items communs aux questionnaires enseignants et parents. Il permet une évaluation par tout observateur extérieur et est particulièrement indiqué pour suivre les enfants sous traitement (Sprague *et al.*, 1973).

Ces questionnaires ont été validés par Goyette *et al.* (1978). Il existe des données normatives selon l'âge et le sexe de l'enfant pour des enfants et adolescents de 3 à 17 ans. Les questionnaires de Conners permettent d'obtenir un profil comportemental assez fiable des enfants, particulièrement dans le registre des comportements externalisés de type hyperactivité, impulsivité ou troubles de conduites. Leur validité est particulièrement bonne en population clinique mais ils peuvent être utilisés en population non clinique.

Il existe une bonne corrélation globale entre les scores d'hyperactivité et l'existence d'un trouble déficitaire de l'attention avec hyperactivité. Plus récemment, d'autres applications ont été trouvées à ces questionnaires, notamment dans l'évaluation des comportements externalisés dans d'autres pathologies comme l'autisme infantile, ou dans des études de dépistage en population générale. Ces questionnaires ont été traduits en français par Dugas et Cook (Dugas, 1987).

• *Exemple d'utilisation des questionnaires de Conners en population générale*

Nous avons mené une enquête auprès des médecins installés en libéral sur les profils psychocomportementaux des enfants allergiques. 1900 médecins ont participé à cette enquête (médecins généralistes, pédiatres, ORL), ce qui a permis d'obtenir une évaluation de 7700 enfants suivis régulièrement pour allergie. Les enfants ont été évalués à l'aide des questionnaires de Conners-parents, complétés par un questionnaire général sur la maladie allergique, le type de traitement et l'adaptation des enfants.

La population était constituée de 7700 enfants, d'âge moyen 8,5 ans, comprenant 2/3 de garçons, 1/3 de filles. Les résultats globaux sont présentés dans le tableau 2.

Les données présentées dans ce tableau confirment la fréquence de difficultés d'ordre psychocomportemental dans cette population et ce de façon plus élevée par rapport à d'autres données établies en population générale (Goyette *et al.*, 1978). Ces altérations du fonctionnement sont retrouvées dans les différents facteurs du questionnaire de Conners-parents mais prédominent dans 4 domaines; les troubles des conduites, les problèmes d'apprentissage, les manifestations psychosomatiques et l'index d'hyperactivité.

Tableau 2. *Fréquence des sous-scores élevés (*T *score* >70*) à l'échelle de Conners. Résultats sur la totalité de l'échantillon*

Facteurs	*N**	*%*
Troubles des conduites	645	9,3
Problèmes d'apprentissage	1185	16,5
Manifestations psychosomatiques	1382	19,6
Anxiété	492	7,0
Impulsif-hyperactif	524	7,6
Index d'hyperactivité	946	13,8

Le tableau 3 présente les résultats du questionnaire de Conners en fonction du type d'allergie. Globalement, il existe des difficultés psychocomportementales plus élevées en cas d'allergie mixte par rapport aux autres allergies isolées.

De plus, les allergies bronchiques isolées entraînent plus fréquemment des perturbations comportementales, particulièrement de type externalisé (troubles des conduites, hyperactivité), que les autres troubles. Cela quel que soit le type de traitement utilisé.

Tableau 3. *Fréquence des sous-scores élevés (*T *score* >70*) à l'échelle de Conners. Résultats par type d'allergie*

	Allergie				
Facteurs	*Cutanée*	*ORL*	*Bronchique*	*Mixte*	*p*
Troubles des conduites	8,5	7,3	10,8	9,5	0,08
Problèmes d'apprentissage	13,2	14	15,1	17,8	0,002
Manifestations psychosomatiques	16,4	16,2	16,9	21,5	0,001
Anxiété	5,2	5,7	5,3	7,9	0,005
Impulsif-hyperactif	5,5	6,2	6,4	8,5	0,02
Index d'hyperactivité	9,7	10,3	13,1	15,4	0,001

En conclusion, les évaluations des perturbations comportementales de l'enfant peuvent être aisément réalisées à l'aide d'instruments simples, d'emploi facile, et aujourd'hui, bien validés. Ces instruments peuvent être utilisés aussi bien dans une évaluation individuelle permettant la recherche d'un profil comportemental particulier chez un enfant que dans des évaluations épidémiologiques recherchant en population générale ou en population clinique la fréquence d'apparition de tel ou tel trouble comportemental, comme nous avons tenté de l'illustrer. Ce type d'évaluation reste naturellement tout à fait complémentaire d'une approche individuelle plus clinique incluant dans la compréhension des troubles l'évaluation du fonctionnement familial et de l'histoire de l'enfant.

ANNEXES

Questionnaire de Conners pour les parents (version à 48 items)

DATE..
NOM ET PRÉNOM DE L'ENFANT .... ÂGE ....
Votre nom.. Lien de parenté à l'enfant ....
La période dont il faut tenir compte pour répondre au questionnaire est la période actuelle.

Vous trouverez ci-dessous des indications décrivant des comportements d'enfants ou les problèmes qu'ils ont parfois. Lisez chaque indication pendant la période indiquée ci-dessus : PAS DU TOUT, UN PETIT PEU, BEAUCOUP, ÉNORMÉMENT

	PAS DU TOUT	UN PETIT PEU	BEAUCOUP	ÉNORMÉMENT
1. Tripote ou ronge certaines choses (ongles, doigts, cheveux, vêtements).				
2. Insolent avec les grandes personnes.				
3. A du mal à se faire des amis et à les garder.				
4. Excitable, impulsif.				
5. Veut tout commander.				
6. Suce ou mâchonne (pouce, vêtements, couvertures).				
7. Pleure facilement ou souvent.				
8. Se sent attaqué, est sur la défensive.				
9. Rêvasse.				
10. A des difficultés d'apprentissage.				
11. Se « tortille », ne tient pas en place.				
12. A peur (de nouvelles situations, endroits, et personnes nouvelles, d'aller à l'école).				
13. Agité, a toujours besoin de faire quelque chose.				
14. Destructeur.				
15. Ment ou raconte des histoires qui ne sont pas vraies.				
16. Timide.				
17. S'attire plus d'ennuis (se fait plus attraper) que les autres enfants de son âge.				
18. Ne parle pas comme les autres enfants de son âge (parle comme un bébé, bégaye, est difficile à comprendre).				
19. Nie ses erreurs ou accuse les autres.				
20. Querelleur.				
21. Fait la moue et boude.				
22. Prend des choses qui ne lui appartiennent pas.				
23. Est désobéissant ou obéit à contre-cœur.				

	PAS DU TOUT	UN PETIT PEU	BEAUCOUP	ÉNORMÉMENT
24. S'inquiète plus que les autres (de la maladie, la mort, la solitude).				
25. Ne termine pas ce qu'il (elle) a commencé.				
26. Se sent facilement froissé.				
27. Brutalise ou intimide ses camarades.				
28. Ne peut pas s'arrêter lors d'une activité répétitive.				
29. Cruel.				
30. Comportement « bébé » ou immature (demande qu'on l'aide pour quelque chose qu'il peut faire tout seul, est « collant », a constamment besoin d'être rassuré).				
31. Problème de fixation de l'attention ou distractibilité.				
32. Maux de tête.				
33. Changements d'humeur rapide.				
34. N'obéit pas ou n'aime pas obéir aux règles ou interdits.				
35. Se bagarre constamment.				
36. Ne s'entend pas avec ses frères et sœurs.				
37. Se décourage facilement lorsqu'un effort est nécessaire.				
38. Dérange les autres enfants.				
39. Enfant foncièrement malheureux.				
40. Problèmes d'alimentation (mauvais appétit, se lève après chaque bouchée).				
41. Maux d'estomac.				
42. Problèmes de sommeil (ne peut s'endormir, se réveille trop tôt, se réveille pendant la nuit).				
43. Autres plaintes physiques et douleurs.				
44. Vomissements, nausées.				
45. Se sent « lésé » à la maison.				
46. Se vante, fanfaronne				
47. Se laisse écraser, manipuler par les autres.				
48. Problèmes d'évacuation intestinale (selles molles, va à la selle irrégulièrement, constipation).				

Reproduit grâce à l'amabilité de C. Keith Conners.

Questionnaire de Conners pour les enseignants (version à 28 items)

NOM DE L'ENFANT.................................. NOM DE L'ENSEIGNANT ....
PRÉNOM DE L'ENFANT .... NOM ET ADRESSE DE L'ÉTABLISSEMENT........
ÂGE.......................... DATE........................

Vous trouverez ci-dessous des indications décrivant des comportements. Placez une croix dans la colonne qui décrit le mieux cet enfant. RÉPONDEZ À TOUTES LES QUESTIONS.

	PAS DU TOUT	UN PETIT PEU	BEAUCOUP	ÉNORMÉMENT
1. Agité, se « tortille » sur sa chaise.				
2. Fait des bruits incongrus quand il ne faut pas.				
3. On doit répondre immédiatement à sa demande.				
4. Fait le malin.				
5. Crises de colère et conduites imprévisibles.				
6. Trop sensible à la critique.				
7. Distrait ou attention fluctuante.				
8. Perturbe les autres enfants.				
9. Rêveur.				
10. Fait la moue et boude.				
11. Humeur changeant rapidement et de façon marquée.				
12. Bagarreur.				
13. Attitude soumise face à l'autorité.				
14. Agité, toujours en train d'aller à droite et à gauche.				
15. S'excite facilement, impulsif.				
16. Demande une attention excessive de l'enseignant.				
17. Semble mal accepté par le groupe.				
18. Se laisse mener par les autres enfants.				
19. Est mauvais joueur.				
20. Semble manquer de capacités à entraîner ou mener les autres.				
21. Difficultés à terminer ce qu'il commence.				
22. Puérile et immature.				
23. Nie ses erreurs ou accuse les autres.				
24. A des difficultés à s'entendre avec les autres enfants.				
25. Peu coopérant avec ses camarades.				
26. S'énerve facilement quand il doit faire un effort.				
27. Peu coopérant avec ses camarades de classe.				
28. Difficultés d'apprentissage.				

Reproduit grâce à l'amabilité de C. Keith Conners.

Ne rien inscrire ici

LISTE DE COMPORTEMENTS POUR LES ENFANTS AGÉS DE 4 A 16 ANS

NOM ET PRÉNOM

SEXE

☐ Garçon ☐ Fille

AGE

LIEU DE NAISSANCE (département ou pays)

DATE D'AUJOURD'HUI

Jour ___ Mois ___ Année ___

DATE DE NAISSANCE

Jour ___ Mois ___ Année ___

CLASSE :

NE VA PAS À L'ÉCOLE ☐

Répondez à ce questionnaire en exprimant *votre* impression sur l'enfant, même si d'autres personnes pourraient en avoir une autre. N'hésitez pas à écrire vos commentaires à côté de chaque question.

ACTIVITÉ PROFESSIONNELLE HABITUELLE DES PARENTS, même si la profession n'est pas actuellement exercée *(Soyez aussi précis que possible et indiquez la nature de l'emploi occupé. Par exemple : employé de mairie, technicien en électricité, professeur de collège, artisan plombier, etc...)*

POUR LE PÈRE : ___

POUR LA MÈRE : ___

QUESTIONNAIRE REMPLI PAR :

☐ Mère

☐ Père

☐ Autre (préciser) :

I. Veuillez indiquer les sports auxquels votre enfant aime le plus participer.
Par exemple : natation, foot-ball, bicyclette, etc...

☐ Aucun

	Par rapport aux autres enfants de son âge, combien de temps y passe-t-il (elle) environ ?				Par rapport aux autres enfants de son âge, comment réussit-il (elle) dans chacun d'eux ?			
	Ne sait pas	Moins que la moyenne	Comme la moyenne	Plus que la moyenne	Ne sait pas	Moins bien	Aussi bien	Mieux
a. ___	☐	☐	☐	☐	☐	☐	☐	☐
b. ___	☐	☐	☐	☐	☐	☐	☐	☐
c. ___	☐	☐	☐	☐	☐	☐	☐	☐

II. Veuillez indiquer les passe-temps, activités et jeux favoris de votre enfant, autres que les sports.
Par exemple : chant, danse, collections, lecture, poupée, piano, etc... (Ne pas compter la télévision ni la radio).

☐ Aucun

	Par rapport aux autres enfants de son âge, combien de temps y passe-t-il (elle) environ ?				Par rapport aux autres enfants de son âge, comment réussit-il (elle) dans chacun d'eux ?			
	Ne sait pas	Moins que la moyenne	Comme la moyenne	Plus que la moyenne	Ne sait pas	Moins bien	Aussi bien	Mieux
a. ___	☐	☐	☐	☐	☐	☐	☐	☐
b. ___	☐	☐	☐	☐	☐	☐	☐	☐
c. ___	☐	☐	☐	☐	☐	☐	☐	☐

III. Veuillez énumérer les organismes, clubs, équipes ou groupes auxquels votre enfant appartient.
Par exemple : scouts, associations sportives, clubs d'échecs, troupe de théâtre, etc...

☐ Aucun

	Par rapport aux autres enfants de son âge, comment y participe-t-il (elle) ?			
	Ne sait pas	Moins activement	Aussi activement	Plus activement
a. ___	☐	☐	☐	☐
b. ___	☐	☐	☐	☐
c. ___	☐	☐	☐	☐

IV. Veuillez indiquer les petits travaux ou tâches que votre enfant fait.
Par exemple : gardes d'enfants, faire les courses, faire la vaisselle, ranger sa chambre, faire son lit, etc...

☐ Aucun

	Par rapport aux autres enfants de son âge, comment s'en sort-t-il (elle) ?			
	Ne sait pas	Moins bien	Aussi bien	Mieux
a. ___	☐	☐	☐	☐
b. ___	☐	☐	☐	☐
c. ___	☐	☐	☐	☐

V. 1. **Combien de très bons amis ou d'amis intimes votre enfant a-t-il environ ?** ☐ Aucun ☐ 1 ☐ 2 ou 3 ☐ 4 ou plus
(Ne pas compter les frères et sœurs)

2. **A peu près combien de fois par semaine font-ils des choses ensemble en dehors de l'école ?** ☐ moins que 1 ☐ 1 ou 2 ☐ 3 ou plus
(Ne pas compter les frères et sœurs)

VI. Par rapport aux autres enfants de son âge, comment votre enfant :

	Moins bien	A peu près pareil	Mieux	
a. S'entend-il avec ses frères et sœurs ?	☐	☐	☐	☐ N'a pas de frères ou sœurs
b. S'entend-il avec les autres enfants ?	☐	☐	☐	
c. Se comporte-t-il envers ses parents ?	☐	☐	☐	
d. Joue-t-il et travaille-t-il tout seul ?	☐	☐	☐	

VII. 1. **Pour les enfants de 6 ans et plus — résultats dans les matières scolaires principales :**

	Insuffisant	En-dessous de la moyenne	Dans la moyenne	Au-dessus de la moyenne
a. Lecture, Français, ou Littérature	☐	☐	☐	☐
b. Histoire, Géographie	☐	☐	☐	☐
c. Calcul ou Mathématiques	☐	☐	☐	☐
d. Sciences	☐	☐	☐	☐
e. ______	☐	☐	☐	☐
f. ______	☐	☐	☐	☐
g. ______	☐	☐	☐	☐

Autres matières scolaires - par exemple : langues étrangères, économie, informatique, etc... Ne pas compter éducation physique, travaux manuels, dessin, musique, etc...

---

2. **Est-ce que votre enfant est dans une classe spéciale ou une école spéciale ?** ☐ Non ☐ Oui - de quel type ?

---

3. **Est-ce que votre enfant a redoublé une classe ?** ☐ Non ☐ Oui - classe(s) et raisons :

---

4. **Votre enfant a-t-il eu des difficultés d'apprentissage ou d'autres problèmes à l'école ?** ☐ Non ☐ Oui - décrivez :

Quand ces problèmes ont-ils commencé ?

Ces problèmes sont-ils terminés ? ☐ Non ☐ Oui - depuis quand ?

Est-ce que votre enfant a une maladie, ou un handicap ? ☐ Non ☐ Oui - décrivez :

Qu'est-ce qui vous préoccupe le plus à son sujet ?

Indiquez les aspects les plus positifs de votre enfant :

Voici une liste de descriptions qui concernent les enfants. Pour chaque item qui s'applique à votre enfant **maintenant** ou **au cours des 2 derniers mois**, entourez le **2** si l'item est **très vrai** ou **souvent vrai** pour votre enfant. Entourez le **1** si l'item est **à peu près vrai** ou **parfois vrai** pour votre enfant. Si l'item n'est **pas vrai** pour votre enfant, entourez le **0.** Répondez à chaque question du mieux que vous pouvez, même si certaines questions paraissent ne pas s'appliquer à votre enfant.

0 = Pas vrai (à votre connaissance) **1 = A peu près vrai, ou Parfois vrai** **2 = Très vrai ou Souvent vrai**

0 1 2 1. A des comportements trop jeunes pour son âge
0 1 2 2. Allergie *(décrivez)* : ____________

____

0 1 2 3. Conteste ou contredit souvent
0 1 2 4. Asthme

0 1 2 5. Se comporte comme quelqu'un du sexe opposé
0 1 2 6. Fait caca dans sa culotte

0 1 2 7. Se vante, est prétentieux (se)
0 1 2 8. Ne peut pas se concentrer ou maintenir son attention longtemps

0 1 2 9. Ne peut pas se débarrasser de certaines pensées; est obsédé(e) par certaines idées *(décrivez)* :

____

____

0 1 2 10. Ne peut pas rester assis(e) tranquille, remue beaucoup, toujours en train de bouger

0 1 2 11. S'accroche aux adultes ou est trop dépendant(e)
0 1 2 12. Se plaint de se sentir seul(e)

0 1 2 13. Embrouillé(e), confus(e)
0 1 2 14. Pleure souvent

0 1 2 15. Cruel (le) avec les animaux
0 1 2 16. Cruel (le), dominateur (trice), méchant(e) envers les autres

0 1 2 17. Rêvasse ou semble perdu(e) dans ses pensées
0 1 2 18. Se fait mal délibérément ou a fait des tentatives de suicide

0 1 2 19. Demande beaucoup d'attention
0 1 2 20. Détruit ses affaires personnelles

0 1 2 21. Détruit des choses appartenant à sa famille ou à d'autres enfants
0 1 2 22. Désobéissant(e) à la maison

0 1 2 23. Désobéissant(e) à l'école
0 1 2 24. Ne mange pas bien

0 1 2 25. Ne s'entend pas bien avec les autres enfants
0 1 2 26. Ne semble pas se sentir coupable après s'être mal conduit(e)

0 1 2 27. Facilement jaloux (se)
0 1 2 28. Mange ou boit des choses non comestibles *(décrivez)* : ____________

____

0 1 2 29. A peur de certains animaux, de situations ou d'endroits autres que l'école *(décrivez)* : ____________

____

0 1 2 30. A peur d'aller à l'école

0 1 2 31. Craint de penser ou faire quelque chose de mal

0 1 2 32. Pense qu'il (elle) devrait être parfait(e)
0 1 2 33. Pense ou se plaint que personne ne l'aime

0 1 2 34. Pense que les autres lui veulent du mal
0 1 2 35. Se trouve bon (ne) à rien ou inférieur(e)

0 1 2 36. Se fait souvent mal, a tendance à avoir des accidents
0 1 2 37. Se bagarre souvent

0 1 2 38. Se fait souvent taquiner, est l'objet de moqueries
0 1 2 39. A de mauvaises fréquentations

0 1 2 40. Entend des bruits ou des voix qui n'existent pas *(décrivez)* : ____________

____

0 1 2 41. Est impulsif (ve) ou agit sans réfléchir

0 1 2 42. Aime être seul(e)
0 1 2 43. Ment ou triche

0 1 2 44. Se ronge les ongles
0 1 2 45. Nerveux (se) ou tendu(e)

0 1 2 46. A des mouvements nerveux ou des tics *(décrivez)* :

____

____

0 1 2 47. Fait des cauchemars

0 1 2 48. N'est pas aimé(e) par les autres enfants
0 1 2 49. Est constipé(e), ne va pas à la selle

0 1 2 50. Trop peureux (se) ou anxieux (se)
0 1 2 51. A des vertiges

0 1 2 52. Se sent facilement coupable
0 1 2 53. Mange trop

0 1 2 54. Surexcité(e) de fatigue
0 1 2 55. Est trop gros (se)

56. Problèmes physiques sans cause médicale connue :
0 1 2 a. Douleurs diverses
0 1 2 b. Maux de tête
0 1 2 c. Nausées, envies de vomir
0 1 2 d. Problèmes avec les yeux *(décrivez)* :

____

0 1 2 e. Problèmes de peau
0 1 2 f. Mal au ventre ou crampes
0 1 2 g. Vomissements
0 1 2 h. Autres *(décrivez)* : ____________

____

0 = Pas vrai (à votre connaissance) **1 = A peu près vrai, ou Parfois vrai** **2 = Très vrai ou Souvent vrai**

0 1 2 57. Frappe ou agresse physiquement les autres
0 1 2 58. Se met les doigts dans le nez, s'arrache les peaux ou se gratte d'autres parties du corps *(décrivez)* : ____

0 1 2 59. Joue avec son sexe en public
0 1 2 60. Joue trop avec ses parties sexuelles

0 1 2 61. A de mauvais résultats scolaires
0 1 2 62. Est maladroit(e) ou a des gestes mal coordonnés

0 1 2 63. Préfère jouer avec des enfants plus âgés
0 1 2 64. Préfère jouer avec des enfants plus jeunes

0 1 2 65. Refuse de parler
0 1 2 66. Ne peut pas s'empêcher de répéter certains actes ; a des « manies » *(décrivez)* : ____

0 1 2 67. Fugue de la maison
0 1 2 68. Crie beaucoup

0 1 2 69. Secret(e), garde les choses pour soi
0 1 2 70. Voit des choses qui n'existent pas *(décrivez)* : ____

0 1 2 71. Manque de naturel ou facilement mal à l'aise
0 1 2 72. Met le feu

0 1 2 73. Problèmes sexuels *(décrivez)* : ____

0 1 2 74. Fait son intéressant(e), se donne en spectacle

0 1 2 75. Timide ou réservé(e)
0 1 2 76. Dort moins que la plupart des enfants

0 1 2 77. Dort plus que la plupart des enfants durant le jour et/ou la nuit *(décrivez)* : ____

0 1 2 78. Se barbouille ou joue avec ses selles

0 1 2 79. Problèmes d'élocution ou de prononciation *(décrivez)* : ____
0 1 2 80. A le regard vide, sans expression

0 1 2 81. Vole à la maison
0 1 2 82. Vole en dehors de la maison

0 1 2 83. Accumule des choses dont il (elle) n'a pas besoin *(décrivez)* : ____

0 1 2 84. A un comportement étrange *(décrivez)* : ____

0 1 2 85. A des idées bizarres *(décrivez)* : ____

0 1 2 86. Est buté(e), désagréable ou irritable

0 1 2 87. A des sautes d'humeur, est lunatique
0 1 2 88. Boude beaucoup

0 1 2 89. Méfiant(e)
0 1 2 90. Dit des gros mots ou des obscénités

0 1 2 91. Parle de se tuer
0 1 2 92. Parle ou marche durant son sommeil *(décrivez)* : ____

0 1 2 93. Parle trop
0 1 2 94. Asticote les autres, souvent en train de les agacer

0 1 2 95. Fait des colères ou s'emporte facilement
0 1 2 96. Pense trop aux questions sexuelles

0 1 2 97. Menace les gens
0 1 2 98. Suce son pouce ou ses doigts

0 1 2 99. Trop préoccupé(e) par l'ordre ou la propreté
0 1 2 100. Difficultés pour dormir *(décrivez)* : ____

0 1 2 101. Fait l'école buissonnière, manque l'école
0 1 2 102. Est lent(e), ou manque d'énergie ou de tonus

0 1 2 103. Malheureux (se), triste ou déprimé(e)
0 1 2 104. Fait beaucoup de bruit

0 1 2 105. Prend de l'alcool ou des drogues *(décrivez)* : ____

0 1 2 106. Actes de vandalisme

0 1 2 107. Fait pipi dans sa culotte
0 1 2 108. Fait pipi au lit la nuit

0 1 2 109. Pleurnichard(e)
0 1 2 110. Voudrait être de l'autre sexe

0 1 2 111. Replié(e) sur soi, ne se mêle pas aux autres
0 1 2 112. S'inquiète, se fait du souci

113. S'il vous plait, préciser ici tout problème de votre enfant qui n'a pas été évoqué ci-dessus :
0 1 2 ____
0 1 2 ____
0 1 2 ____

VÉRIFIER QUE VOUS AVEZ RÉPONDU A CHAQUE QUESTION *SOULIGNER LES QUESTIONS QUI VOUS PRÉOCCUPENT*

Bibliographie

Achenbach T. M. — The Child Behavior Profile : I. Boys Aged 6-11. *Journal of Consulting and Clinical Psychology, 46*, 478-488, 1978.

Achenbach T. M., Edelbrock C. S. — The Child Behavior Profile : I. Boys Aged 12-16 and Girls Aged 6-11 and 12-16. *Journal of Consulting and Clinical Psychology, 47*, 223-233, 1979.

Achenbach T. M., Edelbrock C. S. — *Manual for the Child Behavior Checklist and Revised Child Behavior Profile, Burlington, VT.* University of Vermont, Department of Psychiatry, 1983.

Barkley R. A. — *Hyperactive Children : a Handbook for Diagnosis and Treatment.* The Guilford Press, New York, 1981, 1990 (2nd ed.).

Barthélémy C., Lelord G. — *Les échelles d'évaluation clinique en psychiatrie de l'enfant.* Expansion Scientifique Française, 1991.

Conners C. K. — A Teacher Rating Scale for Use in Drug Studies with Children. *Am. J. Psychiatry, 126*, 884-888, 1969.

Conners C. K. — Symptoms Patterns in Hyperkinetic, Neurotic and Normal Children. *Child Development, 41*, 667-682, 1970.

Conners C. K. — Parent and Teacher Rating form for the Assessment of Hyperkinesis in Children. In : Keller P. A., Ritt L. G. (eds), *Innovations in Clinical Practice : Source Book* (vol 1) : Sarrasota Fl Professional Research Exchange Inc., 1982.

Dugas M., Albert E., Halfon O., Nedey-Sayag M. C. — *L'hyperactivité chez l'enfant.* Collection nodules, PUF, Paris, 1987.

Dugas M., Bouvard M. P., Waller-Perotte D. — *Les échelles d'évaluation chez l'enfant.* Éditions Techniques Encycl. Méd. Chir. (Paris) Psychiatrie, 37200-E-50, 8 p., 1994.

Edelbrock C. S., Achenbach T. M. — A Typology of Child. Behavior Patterns : Distribution and Correlates for Disturbed Children Aged 6-16. *J. Abnormal Child Psychology, 8*, 441-470, 1980.

Fish B. — Methodology in Child Psychopharmacology. In : Efron D. H., Cole J. O., Levine J., Wittenborn J. R. (Eds). *Psychopharmacology, a Review of Progress 1957-1967*, PHS Publication n°1836. US Government Printing Office. Washington DC., 989-1001, 1968.

Fombonne E., Chehdan F., Carradec A. M., Achard S., Navarro N., Reis S. — Le *Child Behavior Checklist* : un instrument pour la recherche en psychiatrie de l'enfant. *Psychiatr. et Psychobiol., 3*, 409-418, 1988.

Fombonne E. — The Child Behavior Checklist and the Rutter Parental Questionnaire : a Comparison between Two Screening Instruments. *Psychol. Med., 19*, 777-785, 1989.

Fombonne E. — The Chartres Study : I. Prevalence of Psychiatric Disorders among French School-Aged Children. *Brit. J. Psychiatry, 164*, 69-79, 1994.

Goyette C. H., Conners C. K., Ulrich R. F. — Normative Data on Revised Conners Parent and Teacher Rating Scales, *J. Abnormal Child Psychology, 6*, 221-236, 1978.

Rutter M. — A Children Behavior Questionnaire for Completion by Teacher; Preliminary Findings. *Journal of Child Psychology and Psychiatry, 8*, 1-11, 1967.

Shaffer D., Gould M. S., Brasic J. *et al.* — A children's Global Assessment Scale (CGAS). *Arch. Gen. Psychiatry, 40*, 1228-1231, 1983.

Sprague R. L., Sleator E. K. — Effects of Psychopharmacological Agents on Learning Disorders. *Pediatric Clinics of North America, 20*, 719-735, 1973.

21

ÉVALUATION DU SENTIMENT DE CONTRÔLE PERSONNEL (LIEU DE CONTRÔLE)

par R. DARDENNES et V. LECLERC

La croyance que nous avons ou pas un certain contrôle sur l'issue d'une situation donnée est une notion commune et facilement accessible. Elle a acquis un rôle de premier plan dans la psychologie contemporaine tant sur le plan théorique qu'empirique. Ainsi, de nombreuses expérimentations animales et humaines montrent les effets délétères d'une exposition à des stimuli aversifs incontrôlables ; les sujets ainsi traités présentent ensuite un déficit motivationnel, associatif et émotionnel : ils s'avèrent plus lents à éviter ces stimuli ou à leur échapper alors que cela est possible (situation contrôlable). Connu sous le nom d'impuissance apprise, ce phénomène mis en évidence par Overmier et Seligman (1967) montre les effets d'une forme d'apprentissage, celle qu'il n'existe pas de relation entre les actions propres du sujet et les événements extérieurs. *A contrario*, une possibilité de contrôle, réelle ou illusoire, a des effets positifs et améliore la performance des sujets (Hiroto, 1974)[1] ou leur bien-être (Schultz, 1976)[1].

Ces expériences soulignent l'importance empirique de cette notion de croyance en un contrôle personnel sur les événements (les renforcements de l'environnement). Sur le plan conceptuel, plusieurs théories cognitives de la motivation font de cette croyance l'un des principaux déterminants du comportement : le comportement émis par un individu dans une situation donnée dépend de la valeur du but que le sujet cherche à y atteindre et de l'expectation que ce but sera atteint au moyen de ce comportement. Ainsi, les théories de l'apprentissage social de Bandura (1980) et de Rotter (1954)[1] sont fondées sur de telles expectations. Ces deux auteurs font des expectations de renforcement, plutôt que des renforcements eux-mêmes, les véritables moteurs de l'apprentissage. Cependant, Bandura fait appel à la notion d'efficacité personnelle dans l'autorégulation des comportements, tandis que Rotter s'intéresse à l'origine perçue et attendue du contrôle d'une situation dans le déterminisme des comportements d'un sujet. Constatant que les renforcements n'exercent pas toujours l'effet prévu sur le comportement, ce dernier postule que leurs effets dépendent d'une médiation cognitive : la perception (ou son absence) d'une relation causale entre son propre comportement et le renforcement. Si cette perception dépend pour une large part de la situation spécifique en cause, le modèle de Rotter comprend un principe de généralisation des attentes de contrôle qui se révèle dans les

1. Les références suivies de cet indice sont citées dans Dardennes (1990 et 1994).

situations nouvelles ou ambiguës; cette expectation générale de contrôle des renforcements est à la base du concept de «lieu de contrôle», ou contrôle interne-externe des renforcements[2].

Disposition générale de la personne, ce concept a acquis le statut d'une variable de personnalité : un sujet est jugé interne s'il croit (généralement) que l'obtention d'un renforcement dépend de lui, il est jugé externe s'il croit qu'elle dépend de la chance, ou est sous le contrôle d'autrui. Cette variable permet, selon Rotter, de prédire en partie le comportement d'un sujet dans des situations variées. Ce concept va connaître un grand succès, sans doute du fait de son opérationnalisation sous forme d'un questionnaire publié (Rotter, 1966); la publication de cette échelle va permettre la prolifération d'études sur les différences individuelles en matière de lieu de contrôle et leurs conséquences dans des domaines aussi variés que la réussite universitaire, l'engagement politique ou la psychopathologie. Cette échelle nous servira de point de départ pour mettre en évidence les questions qu'un tel concept peut susciter et exposer les tentatives d'y répondre; avant de poursuivre par l'examen des études sur les relations entre lieu de contrôle et anxiété ou dépression, nous verrons les facteurs socioculturels influents à prendre en compte.

L'ÉCHELLE I-E DE ROTTER (1966)

Rotter avait pour objectif la mise au point d'un outil facile à administrer, faiblement corrélé à une mesure de désirabilité sociale (tendance à épouser des opinions socialement conformes), afin d'évaluer une attente généralisée de contrôle dans des situations très variées. Après un important travail de validation (réduction de 100 à 23 items par analyses factorielles successives, validation externe par la prédiction de comportements effectifs dans 2 études), l'outil final consiste en un questionnaire de type choix forcé de 23 items. Chaque item propose deux explications opposées sur la perception de contrôle et le sujet doit choisir celle qui correspond le mieux à son opinion. À titre d'exemple, l'item 11 est formulé ainsi[3] :

a) Le succès est une question de travail, presque jamais une question de chance.

b) Pour obtenir une bonne situation, il suffit souvent de se trouver au bon endroit au bon moment.

L'échelle est cotée dans le sens de l'externalité (1 point pour les propositions «externes», 0 pour les «internes»). Un score élevé correspond donc à une orientation externe du lieu de contrôle et inversement pour une orientation interne. La sensibilité du lieu de contrôle à de nombreuses variables sociodémographiques (*cf. infra*) ne permet pas d'établir des normes pour la population générale. La plupart des études effectuent une catégorisation des sujets recrutés en internes et externes par rapport à la moyenne du groupe analysé. Pour la version française de l'échelle de Rotter, la moyenne établie par Salehi (1981) sur un groupe de 210 étudiants était de 10,55 (é.t. 3,89), la note maximale (d'externalité) étant de 23.

2. Les références bibliographiques sur ce sujet sont indexées sous le terme «*internal-external control*» (dans la base documentaire Psyclit ainsi que dans le Medline).

3. La version française de l'échelle de Rotter est d'Agathon et Salehi (1978) et figure dans la thèse de Salehi (1981).

LE LIEU DE CONTRÔLE EST-IL UNIDIMENSIONNEL ET GÉNÉRAL?

Cependant, la simplicité de cet outil n'est qu'apparente. En effet, ce questionnaire a été soumis par différents auteurs à des analyses factorielles qui ont mis en doute ses deux caractéristiques fondamentales : son unidimensionnalité et son caractère général. Une revue de 15 analyses factorielles (Coombs et Schroeder, 1988) montre que seule l'analyse factorielle originale, sur 1000 sujets, était unifactorielle, le nombre de facteurs calculé allant de 2 à 9 et que la part de variance expliquée était en moyenne de 11 % pour le premier facteur (interne). Ces résultats ont donné lieu à diverses interprétations quant aux dimensions du lieu de contrôle. Collins (1974)[1] juge ainsi que le lieu de contrôle a au moins deux dimensions ; l'une porte sur la prévisibilité des événements, l'autre sur la nature situationnelle (externe) ou dispositionnelle (interne) de l'événement. De plus, la nature générale du concept a rapidement été mise en pièces. Déçu par les faibles prédictions réalisées par l'échelle de Rotter dans leur domaine d'intérêt, la plupart des auteurs ont cherché à mettre au point des échelles spécifiques d'un domaine — milieu scolaire ou professionnel, attitude face à la maladie, alcoolisme ou exercice physique — afin d'obtenir des prédictions plus fiables dans des situations précises (de nombreux exemples sont présentés dans Lefcourt, 1981). Cette spécification du lieu de contrôle était envisagée par Rotter ; pour ce dernier, les expectations de contrôle dans une situation donnée sont la combinaison d'expectations spécifiques basées sur l'expérience de situations similaires et d'une expectation générale. Rotter reconnaissait le fait que l'évaluation d'expectations spécifiques d'un domaine permettait de meilleures prédictions, le lieu de contrôle (par définition général) s'exprimant surtout dans les situations nouvelles ou ambiguës. Cette mise en cause du caractère unidimensionnel du lieu de contrôle et l'étude de ce concept dans des domaines très variés ont favorisé le développement d'autres outils d'évaluation du « lieu de contrôle ». Le questionnaire de Rotter reste l'échelle générale la plus utilisée dans le monde mais de très nombreuses échelles ont été publiées[4] depuis sa parution. Selon Dubois (1987), cette prolifération a quatre causes :

1) plusieurs auteurs ont voulu préciser la notion de lieu de contrôle en étudiant ce que mesurait l'échelle de Rotter et redéfinir ce concept en modifiant celle-ci (par exemple, Mirels, 1970)[1] ;

2) de manière similaire, la plupart des chercheurs de ce domaine considèrent que la seule définition d'un concept est sa définition opérationnelle et toute tentative de redéfinition théorique donne lieu à un instrument d'évaluation *ad hoc* ;

3) plusieurs outils spécifiques ont été développés en fonction de la population étudiée (enfants, étudiants, alcooliques, etc.) ou d'un domaine particulier (réussite sociale, santé, etc.) (Lefcourt, 1981) ; et enfin

4) la validation de ces questionnaires est souvent négligée.

Parmi la masse de questionnaires disponibles — une cinquantaine au bas mot — deux questionnaires ont des qualités métrologiques satisfaisantes, l'échelle IPC de Levenson (*The Internal, Powerful Others and Chance Scales*), datant de 1972 et disponible dans Levenson (1981), et l'échelle des Sphères de Contrôle (SOC - *Spheres of Control Scale*) de Paulhus (1990 pour la dernière version). La première est

---

4. Les échelles et questionnaires sont les principaux moyens d'évaluer le lieu de contrôle mais ce ne sont pas les seuls, il existe des tests basés sur des scènes à interpréter (sur le principe du TAT) (Dubois, 1987).

une échelle générale multidimensionnelle dont nous avons commencé la validation en langue française (Loas, Dardennes *et al.*, 1994), tandis que la seconde divise le champ du lieu de contrôle en trois domaines.

L'ÉCHELLE IPC DE LEVENSON

Cette échelle multidimensionnelle du lieu de contrôle distingue une externalité imprévisible (le hasard) et une externalité prévisible (les autres tout-puissants). En effet, pour Levenson, une personne qui croit au contrôle des renforcements par des tout-puissants (*powerful others*) peut percevoir suffisamment de régularité dans leurs actions pour tout de même croire qu'il peut obtenir les renforcements voulus par des actions appropriées de sa part. En outre, une personne qui croit au contrôle par la chance devrait avoir un comportement différent de celle qui croit en une absence de contrôle personnel. Trois types d'expectations sont donc évaluées séparément par trois sous-échelles : celles d'un contrôle interne (échelle I), d'un contrôle par d'autres tout-puissants (échelle P) et d'une absence de contrôle (échelle C). Le questionnaire devait répondre aux exigences suivantes : indépendance statistique de ces trois dimensions, favorisée par un format de réponse de type Likert ; formulation des items à la première personne du singulier afin d'évaluer le contrôle personnel du sujet et non son opinion sur le contrôle qu'ont les gens en général ; parallélisme entre les items de chaque sous-échelle (les trois types d'explications sont proposées pour un même thème) ; enfin, absence de corrélation avec une mesure de désirabilité sociale. La version finale comporte 24 items (8 par sous-échelle) dans un format Likert en six points et donne lieu à trois sous-scores (*cf.* annexe).

À titre d'exemple, voici trois items parallèles pour lesquels le sujet doit exprimer son degré d'accord ou de désaccord :

– Source interne (échelle I) : *Quand je fais des plans pour ma vie, je suis presque sûr de toujours les réaliser.*

– Source « tout-puissants » (échelle P) : *Afin que mes projets marchent, je dois être sûr que les gens qui ont du pouvoir sur moi sont d'accord avec ces projets.*

– Source « chance » (échelle C) : *Il n'est pas prudent que je fasse des projets trop loin dans le temps car, au bout du compte, beaucoup de choses sont dues à la chance ou à la malchance.*

Un sous-score élevé indique une forte croyance en un contrôle par la source désignée. Levenson (1981) insiste sur l'indépendance des sous-scores : « Nous ne pouvons pas interpréter un faible score à l'échelle I (interne) comme l'indice que le sujet croit au hasard ; nous pouvons seulement dire qu'il ne se perçoit pas comme déterminant les événements. »

Plusieurs études de validation ont établi la fidélité et la validité interne de cette échelle (résumée dans Levenson, 1981) et Levenson a montré la pertinence de sa conception multidimensionnelle[5] du lieu de contrôle dans une étude sur la perception du comportement parental (Levenson, 1973)[1]. Cette décomposition du lieu de contrôle en trois sources a été confirmée indépendamment (et involontairement) par Nicole Dubois. Cet auteur a mis au point une échelle de lieu de contrôle (spécifique

5. Ces qualités ont suscité le développement d'échelles multidimensionnelles adaptées à des domaines spécifiques comme la santé (l'échelle de Wallston et Wallston [Wallston, 1981]).

au milieu universitaire) pour vérifier, entre autres, l'unidimensionnalité du lieu de contrôle à l'aide d'une méthode statistique multidimensionnelle non métrique. Si les sources de renforcements peuvent bien s'ordonner sur une dimension, leur répartition montre une division du pôle externe de part et d'autre du pôle interne en fonction de la nature aléatoire ou toute-puissante du contrôle externe (Dubois, 1985a, 1985b). La structure factorielle de cette échelle est stable et la fidélité des sous-échelles satisfaisante (Levenson, 1981; Blau, 1984)[6]. Nous avons donc entrepris (Loas, Dardennes *et al.*, 1994) la validation de la version française de l'échelle dont les premiers résultats sont les suivants : après traduction inverse (de la version traduite en français vers l'anglais US) pour s'assurer du respect de la signification des items du questionnaire, celui-ci a été rempli par un échantillon tout-venant de 134 sujets (53 hommes et 81 femmes) d'âge moyen = 31, 22 ans (extrêmes : 18-67 ans).

Parmi ces sujets, 30 (22,5 %) avaient un niveau inférieur au bac, 14 (10,5 %) avaient le baccalauréat, et 90 (67 %) faisaient des études supérieures.

Les analyses statistiques ont montré une consistance interne des sous-échelles satisfaisante (α de Cronbach, I = 0,58, P = 0,71 et C = 0,68) et l'analyse factorielle avec rotation varimax a mis en évidence une solution à 4 facteurs classant correctement 20 des 24 items de l'échelle, ces 4 facteurs expliquant 36 % de la variance totale (les items de la sous-échelle Chance se répartissent sur 2 facteurs différents). Les ANOVA réalisées n'ont pas montré d'influence du sexe et du niveau d'études; de même les corrélations des scores de l'échelle avec l'âge négligeables (respectivement de 0,198 avec le score I, 0,052 avec le score P et 0,093 avec le score C). Comme l'avait déjà noté Levenson, la sous-échelle Chance est significativement corrélée à la sous-échelle Tout-puissants (*Powerful others*) : $r = 0{,}43$, $p < 0{,}001$, tandis qu'elle n'est pas corrélée à la sous-échelle interne, cette dernière ayant une corrélation modérée ($r = 0{,}22$, $p < 0{,}05$) avec la sous-échelle P. Les scores obtenus sur cet échantillon sont les suivants (les scores obtenus par Levenson dans son étude de validation (Levenson, 1972) sont donnés à titre de comparaison) :

	I (Interne)	*P (Powerful Others)*	*C (Chance)*
Score (écart-type)	28,1 (6,3)	14,7 (6,9)	20,5 (7,4)
Levenson 72, Hommes	36,4 (4,6)	18,9 (7,6)	14,5 (7,7)
Levenson 72, Femmes	35,5 (7,4)	14,6 (6,9)	13,4 (9,1)

Il semble exister une moindre internalité de notre échantillon français et une plus grande croyance en une absence totale de contrôle, c'est-à-dire une plus grande influence de la chance sur les événements.

6. Il nous faut toutefois signaler qu'une analyse factorielle confirmatoire (méthode LISREL) n'a pas retrouvé les 3 facteurs I, P, et C régulièrement retrouvés dans les multiples analyses factorielles effectuées. Cependant, cette étude nous paraît sujette à caution : elle a été effectuée sur des sujets âgés dont les auteurs précisent que certains avaient un déficit cognitif; le format original des réponses a été modifié (transformation d'un Likert à 6 classes en un Likert à 3 classes dont une réponse médiane «*je ne sais pas*») et certaines questions étaient jugées non pertinentes par les sujets étudiés (beaucoup ne conduisaient pas et plus aucun ne travaillait) (Shewchuk, 1990).

Cette échelle distingue donc trois sources de contrôle perçu : la suivante sépare trois domaines comportementaux.

L'ÉCHELLE SOC (*Sphere of Control Scale*) DE PAULHUS (1983, 1990)

Cette échelle unidimensionnelle (le lieu de contrôle a deux pôles, interne et externe) distingue trois « aires concentriques » de contrôle : personnel (PC), interpersonnel (IPC), et sociopolitique (SPC). Paulhus considère que ces trois domaines sont conceptuellement différents et qu'un sujet devrait être caractérisé par un profil de contrôle, par exemple une bonne confiance dans ses capacités de réalisation personnelle, une conviction d'avoir peu de contrôle sur ses relations sociales mais celle de pouvoir changer le monde en participant à des actions politiques. Ce questionnaire est donc constitué de 3 sous-échelles de 10 items chacune, explorant ces différents domaines. Le sujet doit exprimer son degré d'accord (format Likert en 7 points) avec des propositions telles que : « *Je peux habituellement obtenir ce que je veux lorsque j'y travaille dur* » (PC). « *Je me fais facilement des amis et je reste avec eux sans difficulté* » (IPC) et « *Le citoyen moyen peut avoir une influence sur les décisions du gouvernement* » (SPC). Ce questionnaire a montré de bonnes qualités métrologiques — unidimensionnalité des sous-échelles, fidélité test-retest supérieure à 0,9 à 4 semaines — et sa structure en 3 facteurs a été confirmée par deux analyses factorielles confirmatoires LISREL indépendantes (Paulhus, 1983 et Parkes, 1988, référencé dans Paulhus, 1990). Plusieurs études de validité ont été menées par l'auteur et ont confirmé l'intérêt de séparer ces trois domaines : l'évaluation fournie par ces trois sous-échelles distingue des comportements significativement différents dans chacune de ces « sphères » selon que le sujet y est interne ou externe (Paulhus, 1983). Toutefois, il existe une corrélation entre les domaines personnel et interpersonnel (Parkes, 1988).

Bien d'autres échelles sont disponibles mais peu ont fait l'objet de validations aussi poussées que les trois questionnaires présentés. Or, l'évaluation du lieu de contrôle et de sa relation avec un autre paramètre nécessite la prise en compte de plusieurs variables socioculturelles. Il est donc important d'utiliser, dans ce genre d'études, des outils éprouvés.

FACTEURS INFLUENÇANT LE DÉVELOPPEMENT DU LIEU DE CONTRÔLE

Plusieurs précautions sont nécessaires afin d'interpréter les scores d'échelles de lieu de contrôle. Cette croyance en un lieu de contrôle interne ou externe s'acquiert en principe par l'expérience ; il est donc prévisible qu'elle soit différente selon les conditions environnementales.

Les effets de l'âge et du sexe sont toujours controversés. Si on observe une progression de l'enfant dans le sens d'un accroissement de l'internalité avec l'âge, celle-ci semble plus le fait des pratiques éducatives que le produit du développement cognitif à proprement parler[7] (Dubois, 1988)[1]. D'autre part, cette croyance ne peut être d'une stabilité absolue à l'âge adulte si l'on considère qu'elle se forme

7. Cet effet normatif serait non spécifique d'un type de scolarisation et plutôt un effet général du système éducatif (en Occident).

avec l'expérience mais tend à se stabiliser avec l'âge, ce que Lachman et Leff (1989)[1] ont vérifié sur une population de personnes âgées durant une période de cinq ans. L'influence du sexe, variable facilement disponible, est l'objet de contrôles quasi systématiques dans les études sur le lieu de contrôle. Leurs résultats se divisent en deux catégories : absence d'influence ou d'externalité légèrement plus élevée dans le sexe féminin (Lefcourt, 1981 ; Dubois, 1987)[8].

L'orientation du lieu de contrôle est fonction du niveau socio-économique des sujets et de nombreuses études montrent que les groupes favorisés sont les plus internes (Dubois, 1987; Lefcourt, 1981 et Levenson, 1981 citent plusieurs travaux démontrant cette relation) et les groupes défavorisés les plus externes, ces groupes différant fortement quant au contrôle effectif qu'ils expérimentent dans leur vie quotidienne et dans l'obtention d'avantages socio-économiques. Il nous faut toutefois signaler que cette explication n'est pas celle de psychologues sociaux comme Dubois (1987) et Beauvois (1984)[1]. Ces auteurs interprètent cette donnée à la lumière d'une approche idéologique des explications psychologiques (la norme d'internalité). Dubois considère ainsi que les différences constatées reflètent les rationalisations propres aux différentes positions sociales[9].

Enfin, cette croyance générale diffère selon la culture des sujets : les Japonais sont plus externes que les Américains tandis que les citoyens de pays anglo-saxons (États-Unis, Australie et Nouvelle-Zélande) sont très proches dans ce domaine (McGinnies, 1974)[1] et que les Indiens sont plus internes encore que les Américains (Chandler, 1981)[1]; cette croyance est acquise car les immigrants ont des scores intermédiaires entre ceux observés dans leur pays d'origine et ceux du pays d'accueil (McGinnies, 1974)[1]. Une étude a comparé les résultats d'un groupe français à ceux de pays européens (Allemagne fédérale et Italie) et n'a pas mis en évidence de différences entre ces pays (Parsons, 1974).

Ces influences de la culture, de l'âge, du sexe et du niveau socio-économique montrent à l'évidence le caractère acquis et le développement progressif dans l'enfance et l'adolescence des croyances en matière de lieu de contrôle. Il était inévitable d'évaluer les conséquences de ces croyances sur l'adaptation du sujet et leurs relations avec des troubles psychopathologiques.

LIEU DE CONTRÔLE ET PSYCHOPATHOLOGIE

L'idée de départ est qu'un individu qui établit des liens entre son comportement et ses conséquences — les renforcements — s'adapte mieux à l'environnement qu'un sujet croyant à un déterminisme indépendant de ses conduites. L'hypothèse initiale de Rotter était que les sujets les plus extrêmes en matière de lieu de contrôle — les plus internes et les plus externes — devaient avoir des troubles importants de

8. Voir aussi Strickland 1980* pour une analyse des différences dans le choix des items selon le sexe à score global identique.

9. Cependant, les recherches entreprises (influence du type d'éducation — libérale ou autoritaire — sur la formation de la norme d'internalité) n'ont pas donné les résultats escomptés. *cf.* DUBOIS N., LE POULTIER F. — Effets du libéralisme sur l'internalité et la clairvoyance normative. In : BEAUVOIS J. L., JOULE R. V., MONTEIL J. M. (éds) — Perspectives cognitives et conduites sociales (IV), 83-94, Delachaux et Niestlé, Neuchâtel, 1993.

l'adaptation (Rotter, 1966) mais elle n'a pas été confirmée (Rotter, 1975)[1] et il a surtout été observé, et recherché, une corrélation linéaire entre une croyance externe et l'intensité de différents troubles psychopathologiques (*cf.*, par exemple, Archer, 1980; Harrow, 1969; Smith, 1971; Levenson, 1973a)[1].

La dépression a fait l'objet de très nombreuses études (Benassi *et al.* (1988) en ont identifié 427 de 1966 et 1986) et la majorité des travaux sur l'orientation du lieu de contrôle dans la dépression est en faveur d'une relation entre externalité et intensité de la dépression (Benassi, 1988). Cette observation se rapproche des premiers travaux de Seligman concernant l'impuissance apprise (Seligman, 1975)[1] car, au départ, cette théorie faisait de l'expectation d'incontrôlabilité un élément explicatif central de la dépression chez l'homme. Ce même type de relation a été retrouvé entre l'externalité d'une part et l'anxiété-trait d'autre part (Archer, 1979a)[1]; cependant, il existe très peu d'études sur des populations cliniques (Dardennes, 1994).

En fait, les études sur des populations psychiatriques variées montrent une relation globale entre externalité et troubles psychopathologiques, avec une différence quantitative plutôt que qualitative entre les différents troubles psychiatriques. Il existe cependant des nuances dans cette relation. Ainsi, une analyse de régression multiple pas-à-pas nous a permis de montrer, sur une population clinique, que les scores de dépression sont mieux prédits par la sous-échelle P de l'échelle de Levenson, tandis que les scores d'anxiété sont liés à la sous-échelle C (Dardennes, 1994). Le type d'externalité serait donc différent entre ces deux entités : prévisible pour la dépression et aléatoire pour l'anxiété. Toutefois, nous n'avons pas pu montrer de différences de scores de lieu de contrôle entre un groupe de sujets anxieux et un autre de déprimés.

L'interprétation faite par la plupart des auteurs de cette relation entre un lieu de contrôle externe et la présence de troubles psychopathologiques est simple : une orientation externe est un facteur de vulnérabilité non spécifique aux troubles psychopathologiques et leur relation est de type linéaire. Mais association ne veut pas dire relation causale et il est plus probable que l'externalité observée est la conséquence des troubles que leur cause. En effet, certaines études expérimentales montrent que ce sont les sujets internes qui réagissent le plus vivement aux échecs ou aux situations stressantes et effondrent leurs performances dans des situations où le contrôle leur échappe. Ainsi, Houston (1972)[1] montre que les sujets internes ne diffèrent pas des externes quant à l'anxiété exprimée (dans un autoquestionnaire) mais qu'ils réagissent physiquement beaucoup plus que les externes dans les situations de stress. De plus, l'étude des relations entre anxiété et lieu de contrôle dans des situations de test ambiguës quant au contrôle perceptible montre que ces relations ne sont pas de type linéaire (Archer, 1979b)[1]. Un dernier élément doit inciter à la prudence lors de l'interprétation des évaluations du lieu de contrôle pour la dépression; Lamont (1972) a montré l'existence d'un biais d'humeur dans la formulation des items de l'échelle de Rotter, les sujets tristes choisissant plus les items en fonction de leur tonalité d'humeur qu'en fonction de leur orientation interne ou externe.

CONCLUSION

Le lieu de contrôle est l'un des thèmes majeurs de la recherche nord-américaine de ces trente dernières années et son opérationnalisation a donné lieu à de multiples

questionnaires. Les trois questionnaires que nous avons exposés ont fait l'objet de nombreux travaux de validation. L'utilisation de questionnaires éprouvés est une obligation pour qui veut étudier le lieu de contrôle en psychopathologie, étant donné la complexité de ses relations avec de nombreuses variables.

BIBLIOGRAPHIE

BANDURA A. — *L'apprentissage social.* Mardaga, Bruxelles, 1980.

BENASSI V. A., SWEENEY P. D., DUFOUR C. L. — Is there a Relation between Locus of Control Orientation and Depression? *Journal of Abnormal Psychology, 97*, 357-367, 1988.

Canadian Journal of Behavioural Science. *Special Issue : The Psychology of Control, 24*, n° 2, 143-266, 1992.

COOMBS W. N. et SCHROEDER H. E. — Generalized Locus of Control : an Analysis of Factor Analytic Data. *Personality and Individual Differences, 9*, 79-85, 1988.

DARDENNES R. M. — *Le lieu de contrôle de Rotter : contexte théorique, évaluation et intérêt potentiel dans la dépression.* Mémoire pour le DES de psychiatrie. Paris, 1990.

DARDENNES R. M. — *Étude comparée de l'orientation du lieu de contrôle chez les déprimés et anxieux.* Mémoire pour le DEA de psychologie clinique et psychopathologie. Université Paris XIII, 1994.

DUBOIS N. — Contribution à l'étude de la dimensionnalité du concept de « Locus Of Control » (LOC). *L'année Psychologique, 85*, 27-40, 1985.

DUBOIS N. — Une échelle française de « Locus Of Control ». *Revue de Psychologie Appliquée, 35*, 215-233, 1985.

DUBOIS N. — *La psychologie du contrôle.* Presses Universitaires de Grenoble, Grenoble, 1987.

FLEMING J. S., SPOONER P. S. — Five Factor Scales for Internal-External Control and their Relations to Measures of Adjustment. *Journal of Clinical Psychology, 41*, 512-517, 1985.

FURNHAM A. — A Content and Correlational Analysis of Seven Locus of Control Scales. *Current Psychological Research & Reviews, 6*, 244-255, 1987.

LAMONT J., BROOKS R. — Mood-Response Bias in the Rotter I-E Scale. *Journal of Clinical Psychology, 29*, 416-417, 1973.

LEFCOURT H. M. (Ed). — *Research with the Locus of Control Construct* : Vol 1 Assessment Methods. Academic Press, New York, 1981.

LEFCOURT H. M. — Durability and Impact of the Locus of Control Construct. *Psychological Bulletin, 112*, 411-414, 1992.

LEVENSON H. — Differentiating among Internality, Powerful Others and Chances. In : LEFCOURT H. M. (Ed). Research with the Locus of Control Construct : Vol 1 Assessment Methods. Academic Press, New York, 15-63, 1981.

LOAS G., DARDENNES R., DHEE-PEROT P., LECLERC V. — Opérationalisation du concept de « lieu de contrôle » : traduction et première étude de validation de l'échelle de contrôle de Levenson (IPC : *The Internal, Powerful Others and Chance Scale*). *Annales Médico-Psychologiques, 152*, 466-469, 1994.

PAULHUS D. L., VAN SELST M. — The Spheres of Control Scale : 10 Years of Research. *Personality and Individual Differences, 11*, 1029-1036, 1990.

ROTTER J. B. — Generalized Expectancies for Internal Versus External Control of Reinforcement. *Psychological Monographs, 80*, (1, whole n° 609), 1966.

SALEHI I. — *Le questionnaire du lieu de contrôle de J. B. Rotter. Sa base théorique et son intérêt.* Thèse pour le Doctorat de 3[e] cycle. Université Paris V, 1981.

SALEHI I. — Le concept de lieu de contrôle : les différences individuelles. *Journal de Thérapie Comportementale, IV*, 1-10, 1982.

SCHEWCHUK R. M., FOELKER G. A. Jr, NIEDEREHE G. — Measuring Locus of Control in Elderly Persons. *International Journal of Aging and Human Development*, *30*, 213-224, 1990.

WALLSTON K. A. — Hocus-Pocus, the Focus isn't Strictly on Locus : Rotter's Social Learning Theory Modified for Health. *Cognitive Therapy and Research*, *16*, 183-199, 1992.

WHITMAN L., DESMOND S. M., PRICE J. H. — Development of a Depression Locus of Control Scale. *Psychological Reports*, *60*, 583-589, 1987.

ANNEXE
Items de l'échelle IPC de Levenson (traduction R. Dardennes et V. Leclerc, 1991)

1. Le fait que je devienne ou non un leader dépend surtout de mes capacités.

2. Ma vie est en grande partie dirigée par des événements dus au hasard.

3. J'ai le sentiment que ce qui arrive dans ma vie est principalement décidé par les gens puissants.

4. Le fait d'avoir ou de ne pas avoir d'accident de voiture dépend surtout de mon habileté au volant.

5. Quand je fais des projets, je suis presque toujours sûr de les réussir.

6. Il n'est souvent pas possible de protéger mes intérêts contre des événements malchanceux.

7. Quand j'obtiens ce que je veux, c'est généralement par chance.

8. Sans l'intervention de personnes bien placées, on ne me donnera pas un poste à responsabilités même si je suis capable de l'occuper.

9. Le nombre d'amis que j'ai dépend de mon degré de gentillesse personnelle.

10. J'ai souvent constaté que ce qui doit arriver arrive.

11. Le déroulement de mon existence est essentiellement sous le contrôle des gens puissants.

12. Le fait d'avoir ou de ne pas avoir d'accident de voiture est avant tout une question de hasard.

13. Les gens comme moi ont très peu de possibilités de défendre leurs intérêts personnels contre ceux de puissants groupes de pression.

14. Pour moi, il n'est pas toujours sage de faire des projets trop longtemps à l'avance car de nombreuses choses s'avèrent être une question de bonne ou de mauvaise fortune.

15. Pour obtenir ce que je veux, il faut que cela plaise aux personnes placées au-dessus de moi.

16. Parvenir ou non à être un leader dépend de la chance que j'aurai d'être à la bonne place au bon moment.

17. Si des gens importants décidaient qu'ils ne m'aimaient pas, alors je ne me ferais probablement pas beaucoup d'amis.

18. Je peux à peu près déterminer ce qui arrivera dans ma vie.

19. En général, je suis capable de protéger mes intérêts personnels.

20. Le fait qu'il m'arrive ou non un accident de voiture dépend surtout des autres conducteurs.

21. Quand j'obtiens ce que je veux, c'est en général parce que j'ai travaillé dur pour cela.

22. Afin que mes projets réussissent, je m'assure qu'ils correspondent aux désirs des personnes qui ont du pouvoir sur moi.

23. Ce sont mes propres actions qui déterminent ma vie.

24. Que j'aie peu ou beaucoup d'amis, c'est avant tout une question de destinée.

22

L'ÉVALUATION DE L'ANHÉDONIE EN PSYCHOPATHOLOGIE

par G. LOAS

INTRODUCTION

L'anhédonie est un néologisme créé par analogie à l'analgésie par Théodule Ribot en 1896 (*Psychologie des sentiments*) et signifiant la perte ou l'absence de la sensibilité au plaisir.

Afin de mieux cerner cette notion d'anhédonie, il convient de préciser la place du plaisir par rapport aux émotions.

La définition des émotions s'avère difficile dans la mesure où les auteurs ne se basent que sur un seul de ses aspects (réponse, stimulus, expérience subjective...). Ainsi Kleinginna et Kleinginna[1] dans une revue publiée en 1981[1] ont recensé 11 catégories de définitions dépendant étroitement de l'aspect de l'émotion considéré de manière prévalente.

Ainsi que le suggère Kirouac (1989), il existe un consensus pour admettre la distinction entre émotions et autres phénomènes affectifs. Les processus affectifs sont des états impliquant des sensations de plaisir/déplaisir ou liés à la tonalité agréable/désagréable. Les émotions sont des états affectifs, donc impliquant les sensations précédentes, caractérisés par un commencement précis, liés à un objet précis, et de durée limitée dans le temps.

L'humeur se différencie de l'émotion par une durée plus longue et présentant un caractère plutôt chronique comparativement à l'émotion d'évolution plutôt aiguë. Le terme de sentiment, comme celui d'émotion, présente une polysémie, il semble préférable de restreindre son utilisation à celui de la coloration affective des contenus conscients.

Le terme de sentiment employé par Ribot (1896) renvoie à celui d'état affectif et à celui d'émotion ; cet auteur utilise d'ailleurs indistinctement sentiment ou émotion à plusieurs reprises dans la *Psychologie des sentiments*. Il est intéressant de noter que le titre de la traduction anglaise de son ouvrage (1897) est le suivant *The psychology of emotions*.

L'émotion présente plusieurs composantes, une composante d'évaluation cognitive des stimulations ou des situations, une composante physiologique d'activation,

1. Seules les principales références figurent dans la bibliographie située à la fin de l'article. Les références suivies de cet indice figurent dans la thèse de G. Loas (1994) et, en majeure partie, dans l'ouvrage de G. Kirouac (note des éds).

une composante d'expression motrice, une composante d'ébauche d'action et de préparation du comportement et une composante subjective, celle de l'état émotionnel (Scherer, 1984)[1].

Si la définition de consensus évoquée plus haut nous permet de considérer le plaisir comme une émotion, une autre méthode consiste à considérer les travaux portant sur le nombre des émotions, notamment ceux portant sur la théorie des émotions discrètes ou différentielles. Selon ces auteurs (Ekman, 1982; Izard, 1971; Panksepp, 1982)[1], il existe des émotions primaires ou de base d'où découlent des émotions secondaires obtenues par combinaisons des premières, à la manière de la palette d'un peintre (*Théorie de la palette*, Scherer, 1984)[1].

Dans la sélection proposée par Ortony *et al.* (1988)[1] des différentes classifications des émotions primaires selon différents auteurs et à laquelle nous avons ajouté la classification de Ribot (1896), il existe une grande diversité des émotions primaires mais nous remarquons que le plaisir est présent dans 9 des 15 classifications mentionnées.

Pour Ribot (1896), c'est la méthode d'observation et notamment la prise en compte de l'évolution ontogénique qui sert de base à la classification des émotions primaires. Les émotions primaires sont celles qui sont irréductibles aux manifestations antérieures et qui apparaissent comme une manifestation nouvelle, les autres constituant les émotions secondaires. Il distingue ainsi comme émotions primaires, la peur, la colère, l'émotion tendre ou sympathie, l'amour propre et l'émotion sexuelle. Le plaisir comme la douleur sont exclus par leur caractère de généralité (il existe de la douleur dans la peur, dans la colère...) et parce qu'ils imprègnent la vie affective en général contrairement aux émotions primitives qui sont adaptées à une fin particulière (instinct de conservation...).

Sans entrer dans le débat sur la nature et le nombre des émotions primaires, nous pouvons émettre les hypothèses que le plaisir est soit une émotion primaire, soit un état affectif particulier pouvant s'associer au même titre que la douleur aux émotions; les deux hypothèses pouvant être vraies simultanément.

L'étude de l'anhédonie a fait l'objet de nombreux travaux, d'une part au sein des troubles psychiatriques (dépressions et schizophrénies essentiellement) et d'autre part dans le cadre de l'exploration de la personnalité. Dans le premier cas, l'anhédonie est considéré comme un état susceptible de suivre l'évolution du trouble psychiatrique et dans le second elle constitue un trait susceptible de stabilité temporelle (Loas et Pierson, 1989).

MESURE DE L'ANHÉDONIE

La mesure de l'anhédonie chez l'homme peut se faire selon deux modalités, l'une évaluant la perte de la capacité à éprouver du plaisir (anhédonie *stricto sensu*), l'autre évaluant les indices liés au plaisir. À partir de ces deux modalités, il existe plusieurs méthodes de mesure selon que les auteurs privilégient une des composantes du plaisir ou de l'anhédonie, c'est-à-dire l'expression motrice, l'activation physiologique ou l'état émotionnel subjectif. Enfin, au sein de chacune de ces méthodes, il est possible de considérer le plaisir pris dans sa globalité ou de ne prendre en compte que certains de ses aspects (physique, social...).

L'ensemble de ces méthodes peut être divisé en deux groupes, d'une part les méthodes basées sur la psychopathologie quantitative et d'autre part les autres méthodes.

Psychopathologie quantitative

Les instruments utilisés sont de deux types, d'une part les instruments généraux et d'autre part les instruments spécifiques.

Instruments généraux

Nous pouvons y distinguer les critères diagnostiques et les échelles non spécifiques.

• *Critères diagnostiques*

L'échelle pour le diagnostic des troubles affectifs et de la schizophrénie (SADS, *Schedule for Affective Disorders and Schizophrenia*, Spitzer et Endicot, 1978)[1] contient les critères diagnostiques du DSM-III (*Diagnostic and Statistical Manual of Mental Disorders*, Third edition, APA, 1980) et des RDC (*Research Diagnostic Criteria*, Spitzer *et al.*, 1980)[1] notamment pour les dépressions mélancoliques et endogènes. Parmi ces critères figurent la perte de plaisir (RDC) et la perte de plaisir pour toutes les activités (DSM-III et RDC). L'anhédonie a été évaluée à partir de ces critères par plusieurs auteurs (Young *et al.*, 1986; Fawcett *et al.*, 1990)[1]. La fidélité de la SADS a été vérifiée pour la version anglaise. La version française est disponible mais les paramètres métrologiques n'ont pas été publiés à ce jour.

L'index discriminant du Michigan (MDI, *Michigan Discriminant Index*, Feinberg et Caroll, 1982)[1] a été mis au point pour différencier les dépressions non endogènes des dépressions endogènes. Parmi les critères figure la perte de plaisir.

L'échelle de dépression endogénomorphe (EDS) est une échelle de dépression mise au point par Liebowitz en 1981[1], à partir du concept de dépression endogénomorphe de Klein (1974)[1], elle contient un item de perte de plaisir pour toutes les activités.

Le MDI et la EDS ont été utilisés par Davidson *et al.* (1984)[1] pour étudier l'anhédonie. Meehl en 1964[1] a élaboré une *checklist* des signes de schizotypie comprenant notamment l'anhédonie. Cette liste contient 25 items évaluant sur une échelle de 1 à 8 l'absence ou la présence des différents signes de schizotypie (ambivalence, désorganisation cognitive...).

• *Échelles non spécifiques*

Le plaisir et l'anhédonie figurent dans un certain nombre d'instruments utilisés dans la dépression, la schizophrénie ou l'évaluation des émotions.

Dans la dépression et les émotions. Parmi les échelles d'émotions, l'échelle d'émotions différentielles d'Izard (Izard, 1971)[1] permet d'évaluer les émotions primaires dont la joie. Cette échelle existe en version française (Ouss *et al.*, 1990) avec des indices métrologiques satisfaisants (validité de construction et concourante). L'anhédonie a été évaluée tantôt à partir du profil des états de l'humeur (*Profile of Mood States*, McNair *et al.*, 1971)[1] et tantôt en utilisant la liste d'adjectifs affectifs,

(*Multiple Affect Adjective Check List*, Zuckerman et Lubin, 1965)[1]. Jouvent *et al.* ont élaboré en 1988 une échelle d'humeur dont la dernière version (Jouvent *et al.*, 1992) comprend 20 items cotés de 0 à 4. Cette échelle permet d'évaluer l'humeur dépressive en distinguant des signes d'observation et des symptômes ressentis par le patient; sur les 20 items, 3 évaluent l'anhédonie. Les différentes versions de l'échelle ont fait l'objet d'analyses en composantes principales sur des groupes de déprimés majeurs (DSM-III) et ont montré le démembrement de l'humeur dépressive en 3 à 5 composantes (anhédonie, hyperémotivité/irritabilité ressentie, hypo/hyper-expressivité émotionnelle, tristesse, anxiété hypervigilante). Comme le signalent les auteurs (Jouvent *et al.*, 1987)[1], les qualités métrologiques requises pour utiliser leur échelle en usage courant ne sont pas encore établies.

L'échelle d'anxiété et de dépression de Zigmond et Snaith[1] (*The Hospital Anxiety and Depression Scale*) a été mise au point par les auteurs en 1983. Ce questionnaire contient 14 items, 7 d'anxiété et 7 d'anhédonie (représentant la mesure de la dépression), chaque item présente quatre degrés (0 à 3), les scores aux 2 sous-échelles variant de 0 à 21. La validité prédictive et la validité interne (corrélation item score total de la sous-échelle) sont satisfaisantes, mais les auteurs n'ont pas effectué de test-retest ni d'analyse de la cohérence interne (coefficient alpha et analyse factorielle). Cet instrument a été traduit et validé en français par Lépine *et al.* (1985)[1].

Parmi les échelles de dépression, Fleminger (1991)[1] a ajouté trois items à l'inventaire de dépression de Beck pour mesurer l'anhédonie.

Dans la schizophrénie. Grinker *et al.*, en 1973[1], ont étudié l'anhédonie chez les schizophrènes à partir d'un inventaire de schizophrénie contenant un item de plaisir. Cet inventaire n'a pas fait l'objet d'étude de validation.

Andreasen, en 1981, a mis au point l'échelle d'évaluation des symptômes déficitaires (*Scale for the Assessment of Negative Symptoms*, SANS) comprenant cinq composantes dont l'anhédonie associée au retrait social. Les raisons du choix de l'anhédonie comme symptôme négatif n'ont pas été précisées par l'auteur et l'association au retrait social présuppose une association entre ces 2 symptômes, ce qui n'est pas neutre sur le plan théorique. Cette échelle a été traduite en français par Lecrubier et Boyer, mais les études de validation n'ont pas été encore publiées (Aubin *et al.*, 1991)[1].

La composante anhédonie-retrait social comporte cinq items : intérêts et activités de loisirs, intérêts et activités sexuels, incapacité à vivre des relations étroites ou intimes, relations avec les amis et collègues, conscience de l'anhédonie et du retrait social (item subjectif), évaluation globale de l'anhédonie et du retrait social. La SANS est une échelle de type Likert avec 5 degrés de réponses.

Instruments spécifiques

En 1970, Watson *et al.* ont construit une échelle d'anhédonie de 27 items extraits du MMPI. Cette échelle corrèle significativement avec les échelles d'anhédonie physique et sociale de Chapman *et al.* (Peterson et Knudson, 1983)[1].

En 1972, Watson *et al.* ont mis au point une échelle d'anhédonie (*Anhedonia Scale*) à partir de l'échelle d'apathie de Gurel (Gurel, 1961)[1] et de l'échelle d'intérêt d'Elgin (Becker, 1956)[1]. L'échelle d'anhédonie est constituée de cinq items : déficit des intérêts, indifférence affective, manque de dynamisme, retrait social et refus des responsabilités. Les items sont évalués par un cotateur extérieur sur la base d'un entretien semi-structuré. Les paramètres métrologiques ont été déterminés chez

76 hommes présentant des pathologies psychiatriques diverses (sans critères diagnostiques).

La validité interne a été étudiée par le coefficient alpha de Cronbach dont la valeur est de 0,76 et la fidélité au changement par un test-retest avec un coefficient de corrélation de 0,82. Cette échelle a surtout été utilisée par ses auteurs. D'autres auteurs ont critiqué le caractère trop indirect de la mesure de l'anhédonie, notamment par ses conséquences sociales (Chapman *et al.*, 1976).

En 1971, Mac Phillamy et Lewinsohn ont élaboré une échelle d'événements agréables comprenant 320 items portant sur des activités diverses. Les activités sont cotées selon leur fréquence et leur caractère subjectif agréable. Une version abrégée à 80 items a été mise au point en 1976 (Lewinsohn *et al.*, 1976)[1]. La validité, la consistance interne et la fidélité de cette échelle ont été rapportées par leurs auteurs (Mac Phillamy et Lewinsohn, 1972, 1973)[1]. Le principal défaut de cet instrument est lié à l'existence d'activités agréables ne pouvant intéresser que certains sujets (activités sociales particulières...) et rendant difficile la mesure de différences individuelles.

En 1976, Chapman *et al.* ont mis au point deux échelles d'anhédonie, l'échelle d'anhédonie physique (*Physical Anhedonia Scale*, PAS) et l'échelle d'anhédonie sociale (*Social Anhedonia Scale*, SAS). Ces échelles sont des questionnaires comprenant 40 items pour la PAS et 48 pour la SAS décrivant des situations hédoniques caractérisant non pas l'état actuel du sujet mais ses caractéristiques au long cours. La PAS a été révisée en 1978 (Chapman et Chapman, 1978)[1] avec augmentation du nombre d'items (61) pour en augmenter la fidélité et la cohérence interne. La SAS a été révisée (40 items) en 1982 (Eckblad *et al.*, 1982)[1] par élimination des items représentant une anxiété sociale au profit d'items en rapport avec le retrait social schizoïde. Ce retrait social se définit par une indifférence sociale et une incapacité permanente à établir des contacts sociaux. La version révisée de la SAS évalue une anhédonie sociale proche de celle de la personnalité schizoïde du DSM-III (Mishlove *et al.*, 1985)[1]. L'élimination des items portant sur une anxiété sociale liée à un événement social ou une hypersensibilité était motivée par la nécessité de ne pas évaluer un retrait social comme celui de la personnalité évitante du DSM-III. Les paramètres métrologiques de ces instruments ont été étudiés de manière itérative sur de nombreux groupes de sujets sains et malades. Ces instruments ont été adoptés par les auteurs travaillant sur l'anhédonie, notamment la PAS qui s'avère la plus utilisée car elle mesure un déficit plus biologique, moins contaminé par d'autres variables.

Brown *et al.* ont mis au point en 1978 et en 1979 deux échelles de plaisir : une échelle d'auto-évaluation du plaisir lié aux activités (*Pleasurable Activity Self-Rating Scale*) et une échelle d'hétéro-évaluation (*Pleasurable Activity Observer Rating Scale*) traduisant dans le comportement observé les items de l'échelle précédente. Ces échelles contiennent 92 items cotés selon 7 degrés. Les paramètres métrologiques n'ont pas été déterminés.

En 1983, Fawcett *et al.* ont élaboré une échelle de plaisir (*Fawcett Clark Pleasure Capacity Scale*, FCPCS) destinée à l'étude de l'anhédonie dans la dépression. Ce questionnaire contient 36 items portant sur les aspects physiques et sociaux du plaisir. Chaque item est coté en cinq points. La validité et la fidélité ont été déterminées par les auteurs.

Cette échelle a fait l'objet d'une traduction et d'une validation en langue française par Hardy *et al.* (1986) sur un groupe de 120 sujets sains et de 81 déprimés (DSM-III). Les auteurs ont montré la validité interne de la version française de

l'échelle par une analyse en composantes principales (ACP) et par l'étude des corrélations entre les items et le score total. La validité discriminante était insuffisante, les déprimés n'ayant pas un score significativement différent de celui des témoins. Une sous-échelle de plaisir physique (FCPCS-PP) a été extraite de la version complète en français (Loas et Boyer, 1992).

En 1984, Dworkin et Saczynski ont élaboré trois échelles d'anhédonie à partir de l'inventaire multiphasique de personnalité du Minnesota (MMPI) et de l'inventaire psychologique de Californie (CPI). La première échelle contient 33 items du MMPI, la seconde 24 du CPI et la troisième 48 items parmi les précédents. Le coefficient alpha a été calculé chez trois groupes de plus de 100 sujets avec des valeurs variant entre 0,67 et 0,89. La fidélité au changement (test-retest) est attestée par des coefficients de corrélation de 0,85 à 0,89. Le CPI et le MMPI sont disponibles en langue française.

En 1986, Stéphanatos *et al.*[1] ont utilisé des échelles visuelles analogiques portant sur le plaisir éprouvé seul ou avec autrui, ces échelles sont dérivées des *"visual mood and alert scales"* de Luria (1979)[1].

Kadzin en 1989 a mis au point un questionnaire de plaisir destiné aux enfants de 6 à 13 ans comprenant 39 items avec trois degrés de réponse. Le coefficient alpha était de 0,96 et l'analyse en composantes principales effectuée permettait de conclure à une seule dimension. La validité prédictive était objectivée par un score plus élevé chez les enfants déprimés par rapport aux témoins. La validation de la version française est en cours (Loas *et al.*).

Autres méthodes

La nécessité de la reproductibilité des travaux explique le nombre limité des méthodes de mesure ne faisant pas appel à la psychopathologie quantitative.

Kayton et Koh en 1975[1] ont utilisé une tâche de rappel de mots pour évaluer l'anhédonie. Miller en 1987[1] a utilisé des jeux vidéo (jeux d'arcades : briques, packman...) pour mesurer la dimension hédonique.

En 1990, Berenbaum *et al.*[1] ont évalué l'anhédonie à partir de jugements *a posteriori* d'enregistrements audio d'entretiens cliniques. Pour l'évaluation de l'anhédonie, les auteurs ont défini une grille de cotation basée notamment sur des items issus de la PAS, de la SAS et de la *checklist* de Meehl (1964)[1]. Les sujets étaient évalués de 1 (pas anhédonique) à 5 (extrêmement anhédonique). La fidélité inter-juges était évaluée par le coefficient intra-classe (0,82).

MISES AU POINT ET ÉTUDES DE VALIDATION DE L'ÉCHELLE DE PLAISIR PHYSIQUE (FCPCS-PP) ET DE LA VERSION FRANÇAISE DE L'ÉCHELLE D'ANHÉDONIE PHYSIQUE DE CHAPMAN (PAS)

Présentation des échelles

Échelle d'anhédonie physique de Chapman (PAS)

Chapman, Chapman et Raulin ont mis au point en 1976 une échelle d'auto-évaluation de l'anhédonie physique, la PAS (*Physical Anhedonia Scale*) comprenant

40 items, le plaisir physique étant défini par le plaisir lié à la nourriture, au toucher, aux odeurs, au sexe, à la température, au mouvement, aux sons, aux sensations physiques. En 1978, Chapman et Chapman ont ajouté 21 items constitués par des affirmations auxquelles le sujet doit répondre par oui ou non (ex : «*On surestime beaucoup la beauté des couchers de soleil*»). Les instructions de passation demandent au sujet de se décrire tel qu'il est habituellement et non seulement selon l'état dans lequel il se trouve au moment de la passation de l'échelle. Dans cette optique, le temps des verbes est souvent le passé. Chaque réponse est cotée 0 ou 1, le score total de l'échelle variant de 0 à 61.

Après avoir obtenu l'autorisation de l'auteur pour traduire et adapter son échelle en français, nous avons dans un premier temps effectué la traduction. Dans un second temps, une retraduction en américain a été effectuée par un psychiatre bilingue d'origine américaine (Loas, 1993).

Échelle de plaisir physique (FCPCS-PP)

Afin de disposer d'une 2e échelle d'évaluation de l'anhédonie physique, nous avons décidé d'extraire de l'échelle de plaisir de Fawcett *et al.* une sous-échelle de plaisir physique, appelée FCPCS-PP (*Fawcett Clark Pleasure Capacity Scale – Physical Pleasure*).

Le niveau d'anhédonie physique sera mesuré de manière inversement proportionnelle par le score à cette échelle.

L'échelle de plaisir de Fawcett comprend 36 items cotés de 1 à 5. Les instructions demandent au sujet d'imaginer le plaisir qu'il éprouverait lors de situations décrites par les items s'il les vivait actuellement. Cette échelle a été traduite et validée en français par Hardy *et al.* (1986), mais la validité discriminante est faible car les déprimés n'ont pas un score significativement plus bas que les témoins. La cotation des items est différente de celle de la version originale avec des valeurs allant de 5 à 9, soit un score total de 180 à 324.

Cette modification était en rapport avec l'adjonction de 36 items de déplaisir cotés de 1 à 5 et réalisant une échelle de déplaisir (*Hardy Displeasure Capacity Scale*, HDCS).

La HDCS et la version française de la FCPCS constituent une échelle de plaisir-déplaisir cotée de 1 à 9. Les cotations de 1 (déplaisir majeur et ultime) à 5 (pas de déplaisir) concernent les 36 items de déplaisir et les cotations de 5 (pas de plaisir) à 9 (plaisir majeur et ultime) concernent les 36 items de plaisir.

Cette présentation particulière de la FCPCS et de la HDCS permet de faire passer ces 2 échelles simultanément.

Nous avons extrait 12 items portant sur les aspects physiques et sensoriels du plaisir. La méthode était la suivante. Un panel d'experts comprenant 4 psychiatres et une psychologue a été chargé d'extraire parmi les 36 items de la FCPCS ceux se rapportant aux plaisirs physiques et sensoriels. La définition du plaisir sensoriel et physique était celle donnée par Chapman *et al.* en 1976 lors de la mise au point de la PAS. Dans un 2e temps, les items sélectionnés par 3 des 4 experts ont été retenus.

L'échelle de plaisir physique appelée FCPCS-PP (*Fawcett Clark Pleasure Capacity Scale, Physical Pleasure*) comprend 12 items avec un score total variant de 60 à 108 (ex : «*Assis, vous regardez un magnifique coucher de soleil dans une région sauvage du monde*») (Loas et Boyer, 1992).

De manière symétrique, 9 items de déplaisir sensoriel et physique ont été retenus à partir de l'échelle de déplaisir (HDCS) réalisant l'échelle de déplaisir physique (*Hardy Displeasure Capacity Scale-Physical Displeasure*, HDCS-PD).

L'association de la FCPCS-PP et de la HDCS-PD constitue l'échelle de plaisir-déplaisir physique (EPD-P ou *Physical Pleasure-Displeasure Scale*, P-PDS).

Étude de validation

Les études de validation ont consisté à déterminer la validité interne, la validité concourante et la validité discriminante. La fidélité temporelle était appréciée par un test-retest.

La validité interne était objectivée d'une part par le coefficient alpha de Cronbach (KR 20) qui est une fonction de la moyenne des intercorrélations des items et d'autre part par la moyenne des corrélations items/score total. Pour la FCPCS-PP, des analyses en composantes principales furent effectuées.

La validité concourante était mesurée par la corrélation entre les 2 instruments alors que la validité discriminante était évaluée par l'existence d'une anhédonie significativement plus importante chez des sujets malades comparativement à des témoins.

Les résultats des études de validation dans différents groupes de sujets sains et malades sont donnés dans le tableau (Loas *et al.*, 1994a, Loas *et al.*, 1994b, Loas *et al.*, 1995).

Tableau 1. *Paramètres métrologiques de la PAS (à gauche) et de la FCPCS-PP (à droite)*
(moy. cor. item/ST : moyenne des corrélations item/score total, ACP : analyse en composantes principales, toutes les valeurs sont significatives avec $p < 0,05$)

Groupe	*KR 20*	*Moy. cor. item/ST*	*ACP*	*Validité concourante*	*Test-Retest*
61 témoins	0,7/0,7	0,28/0,52	non/oui	– 0,52	
61 dép. RDC	0,83/0,78	0,28/0,52	non/oui	– 0,48	
221 témoins	0,75/0,71	0,25/0,48	non/oui	– 0,39	
32 témoins				– 0,39	0,76/0,62
61 sch. RDC	0,82/0,77			– 0,37	

dép = déprimés ; sch = schizophrènes ; RDC = *Research Diagnostic Criteria.*

Normes en population française

Afin de disposer de normes concernant la PAS et la FCPCS-PP, nous avons constitué deux échantillons de sujets sains provenant de milieux socioprofessionnels variés et d'origine géographique différente (région parisienne et province). Les sujets n'avaient pas d'antécédents psychiatriques et présentaient un score inférieur à 10 à l'inventaire abrégé de dépression de Beck.

La PAS a été remplie par 384 sujets (154 hommes et 230 femmes), d'âge moyen 31,76 ans (sd = 12,24, extrêmes de 17 à 76).

Le score moyen à la PAS est de 15,57 (intervalle de confiance à 95 % de 14,91 à 16,22, sd = 6,53 avec des extrêmes de 2 à 34).

Le score à la PAS ne corrèle pas avec l'âge ($r = -0,07$) et l'analyse de variance comparant les sexes montre une tendance à la significativité ($F(1,383) = 3,26$, $p = 0,07$), les hommes étant plus anhédoniques ($m = 16,24$, $sd = 6,4$) que les femmes ($m = 15$, $sd = 6,6$).

La FCPCS-PP a été remplie par 405 sujets (165 hommes, 240 femmes) répondant aux mêmes critères que précédemment. Les sujets avaient un âge moyen de 34,4 ans (sd = 13,33, avec des extrêmes de 17 à 76). Le score à la FCPCS-PP est de 91,57 (intervalle de confiance à 95 % de 90,96 à 92,18, sd = 6,23, avec des extrêmes de 67 à 106).

Le score à la FCPCS-PP corrèle significativement et négativement avec l'âge ($r = -0,2$, $p < 0,001$) et l'analyse de variance montre une différence significative entre les sexes ($F(1,404) = 6,23$, $p = 0,013$). les hommes sont significativement plus anhédoniques ($m = 90,64$, $sd = 6,29$) que les femmes ($m = 92,21$, $sd = 6,13$).

CONCLUSION

L'anhédonie ou perte de la capacité à éprouver du plaisir constitue une dimension importante à considérer en psychopathologie. Parmi les instruments d'évaluation de l'anhédonie disponibles en langue française, deux semblent particulièrement intéressants :

– l'échelle de plaisir-déplaisir de Fawcett et Clark validée en français, surtout utilisée au cours des états dépressifs dont on peut extraire une sous-échelle de plaisir-déplaisir physique ;

– et l'échelle d'anhédonie physique de Chapman, susceptible d'applications variées chez des sujets sains, chez des patients schizophrènes ou déprimés et chez des sujets ayant un trouble de la personnalité.

L'emploi de ces instruments permet l'étude des relations, toujours discutées, entre la capacité hédonique et la capacité à éprouver des émotions négatives. Chez les patients déprimés, il semble que la capacité hédonique physique soit indépendante de la capacité à éprouver des émotions négatives physiques.

L'exploration de ces dimensions en population générale et au sein de divers échantillons psychopathologiques s'avère pleine de promesses.

BIBLIOGRAPHIE

BROWN S. R., SCHWARTZ G. E., SWEENEY D. R. — Dissociation of Self-Reported and Observed Pleasure in Depression. *Psychosomatic Medicine, 40*, 536-548, 1978.

BROWN S. R., SWEENEY D. R., SCHWARTZ G. E. — Differences between Self-Reported and Observed Pleasure in Depression and Schizophenia. *Journal of Nervous and Mental Disease, 167*, 410-415, 1979.

CHAPMAN L. J., CHAPMAN J. P., RAULIN M. L. — Scales for Physical and Social Anhedonia. *Journal of Abnormal Psychology, 85*, 4, 374-382, 1976.

DWORKIN R. H., SACZYNSKI K. — Individual Differences in Hedonic Capacity. *Journal of Personality Assessment, 48*, 620-626, 1984.

FAWCETT J., CLARK D. C., SCHEFTNER W. A., GIBBONS R. D. — Assessing Anhedonia in Psychiatric Patients : the Pleasure Scale. *Arch. Gen. Psychiatry, 40*, 79-84, 1983.

HARDY P., JOUVENT R., LANCRENON S., ROUMENGOUS V., FELINE A. — L'échelle de plaisir-déplaisir : utilisation dans l'évaluation de la maladie dépressive. *L'Encéphale, XII*, 149-154, 1986.

JOUVENT R., CARTON S., WIDLÖCHER D. — Sémiologie complexe des émotions. *Neuropsychiatrie, 7*, 5, 223-227, 1992.

KAZDIN A. E. — Evaluation of the Pleasure Scale in the Assessment of Anhedonia in Children. *Journal of the American Academy of Child and Adolescent Psychiatry, 7*, 213-221, 1989.

KIROUAC G. — *Les émotions. Monographies de psychologie, 8*, Presses de l'Université du Québec, Québec, 1989.

LOAS G., PIERSON A. — L'anhédonie en psychiatrie. *Ann. Médico-Psychol., 147*, 7, 705-717, 1989.

LOAS G., BOYER P. — Anhedonia and Major Depressive Disorder. In : *New Pharmacological Approaches in the Therapy of Depressive Disorders*. International Academy for Biomedical and Drug Research, Bruxelles, 1992.

LOAS G. — Adaptation et validation française de l'échelle d'anhédonie physique de Chapman et Chapman. *L'Encéphale, XIX*, 639-644, 1993.

LOAS G. — L'anhédonie physique en psychopathologie. Thèse de Science, Paris VI, 1994.

LOAS G., BOYER P. *et al.* — L'évaluation de l'anhédonie en psychopathologie : seconde étude de validation de la version française de l'échelle d'anhédonie physique de Chapman et Chapman (PAS) : étude portant sur 356 personnes. *Ann. Médico-Psychol., 152*, 3, 256-259, 1994a.

LOAS G., SALINAS E., PIERSON A., GUELFI J. D., SAMUEL-LAJEUNESSE B. — Anhedonia and Blunded Affect in Major Depressive Disorder. *Comprehensive Psychiatry, 35*, 366-372, 1994b.

LOAS G., BOYER P., FREMAUX D. *et al.* — L'échelle de plaisir-déplaisir physique (EPD-P) : étude de validation chez 295 sujets, *Rev. Eur. Psychol. Appl., 45*, 139-144, 1995.

MAC PHILLAMY D. J., LEWINSOHN P. M. — The Pleasant Schedule. University of Oregon, Eugene, 1971. Traduction française par WILLMOTTE J. *et al.* disponible dans : COTTRAUX J., BOUVARD M., LEGERON P. — *Méthodes et échelles d'évaluation des comportements*. EAP, Issy-les-Moulineaux, 1985.

OUSS L., CARTON S., JOUVENT R., WIDLÖCHER D. — Traduction et validation de l'échelle d'émotions différentielles d'Izard. *L'Encéphale, XVI*, 453-458, 1990.

RIBOT T. — *La psychologie des sentiments*. Félix Alcan, Paris, 1896.

WATSON C. G., KLEIT W. G., LOREI T. W. — Toward an Operational Definition of Anhedonia. *Psychological Reports, 26*, 371-376, 1970.

23

INSTRUMENTS DE MESURE DE L'ADAPTATION SOCIALE

par P. KOWAL, J. GUELFI et V. GAILLAC

INTRODUCTION

La santé a été définie par l'OMS comme « un état de bien-être physique, mental et social ». Il ne s'agit plus seulement de traiter les symptômes de la maladie mais également de réintégrer le sujet dans sa vie sociale. Cette extension de la médecine vers la médecine ou ici la psychiatrie sociale provient de la réflexion d'auteurs sur les liens entre maladie et insertion sociale. Durkheim (Lewis, 1953) avait déjà relevé que la perte de relations sociales avait un effet négatif sur le bien-être psychologique. La maladie peut conduire à une maladaptation sociale et, à l'inverse, l'environnement social peut conduire à la maladie. Comment définir, partant objectiver et mesurer l'adaptation sociale? Elle peut être comprise comme un ajustement entre le sujet et son entourage. Ceci peut être évalué en fonction soit de la façon dont le sujet se moule dans son rôle social selon les attentes culturelles, soit encore du degré de satisfaction du sujet dans son insertion sociale. Les concepts abondent et aboutissent à bâtir des instruments de mesure nombreux et qui ne se recoupent pas. Nous allons présenter quelques-uns de ces instruments ainsi que la dernière version du DSM en ce qui concerne ce sujet.

INSTRUMENTS DE MESURE DE L'ADAPTATION SOCIALE

Katz Adjustment Scale (KAS) (Katz, 1963)

La KAS est une des plus anciennes et est fréquemment utilisée. Son but original était de mesurer l'ajustement social des patients psychiatriques suivant un traitement. Les auteurs ont considéré que l'ajustement représente un accord entre soi et son environnement. Cette conception se traduit par une étude des sentiments de bien-être et de satisfaction du sujet sur ces performances sociales, en prenant aussi en compte le jugement des proches sur son niveau de fonctionnement. La KAS se présente sous la forme de deux groupes de cinq échelles. Le premier groupe est rempli par le sujet (S scales de S1 à S5), le deuxième par un proche (R scales, R pour relative, de R1 à R5). Pour R1 (127 items), le proche cote les symptômes psychiatriques du patient ainsi que son comportement quand il est en situation sociale (pleure sans raison, recherche le conflit avec autrui...). La fréquence des troubles est notée sur une

échelle en 4 points. R2 décrit les performances dans 16 activités sociales. Une échelle en 3 points est utilisée. R3 concerne les attentes sur le niveau de performance du patient dans les activités de R2. R4 mesure les performances dans 22 activités de loisir. R5 comme R3 concerne les attentes sur les performances de R4. S1 contient 55 items dérivés de la *Hopkins Symptom Check-list* sur l'humeur et les symptômes somatiques. Les items de S2 et suivantes sont équivalents à ceux de R2 et suivantes. La consistance interne de R1 a été trouvée en moyenne égale à 0,72. L'échelle différencie de façon significative les patients qui ont une bonne adaptation de ceux qui en ont une mauvaise. Une analyse factorielle a identifié trois facteurs : adaptation sociale chaotique, psychotisme aigu et retrait/dépression. Il faut 45 à 60 minutes pour la remplir. Cet instrument a souvent été utilisé. Avec le temps, les groupes S ont perdu leur importance et les groupes R ont été privilégiés. C'est pourtant cette approche originale de comparer le point de vue des sujets et de leur proche qui a inspiré d'autres auteurs. La KAS n'évalue pas le sujet dans un rôle social spécifique mais dans des activités ce qui limite son utilisation sur une population de personnes âgées. La validité et la fiabilité de cet instrument ont été l'objet de peu d'études et son utilisation doit être réservée aux cas où les échelles plus actuelles ne sont pas indiquées.

The Social Maladjustment Scale (SMS) (Clare, 1978)

Cet instrument s'est inspiré du précédent à la fois quant à la population — initialement des adultes présentant des troubles psychiatriques plus particulièrement névrotiques — et pour la double cotation par les proches et le patient. La mesure veut ne pas tenir compte de la conformité au rôle social pour ne pas grever d'éventuelles comparaisons interculturelles. Le degré de dysfonctionnement social est mis en valeur. Le SMS se présente sous la forme d'un entretien de 26 pages qui couvrent 6 domaines : la maison, le rôle social ou occupationnel, la situation économique, les loisirs et les activités sociales, les relations familiales et enfin les relations matrimoniales. Les questions couvrent trois thèmes : ce que le sujet a, ce qu'il fait avec (comment il organise sa vie sociale dans ces domaines) et s'il en est satisfait. L'entretien dure 45 minutes et il est réalisé par un cotateur formé au manuel. Un questionnaire de 41 items a été développé à partir du SMS. La fiabilité intercotateurs est bonne. La validité du score global comparée à un entretien standardisé de Goldberg a été calculée avec une sensibilité de 30 % et une spécificité de 80 %. Cette échelle n'a pas été largement utilisée mais elle trouverait sa place lors de l'étude de la maladaptation sociale sous réserve de revoir les critères statistiques ci-dessus.

The Social Relationship Scale (SRS) (McFarlane, 1981)

Cette échelle a été développée pour mesurer l'étendue du réseau social du sujet et la perception qu'il a de son aide effective au cours des événements de sa vie. Elle couvre 6 domaines susceptibles d'évoluer avec le temps : le travail, les finances, la santé, les événements familiaux, sociaux personnels et de société. Sur chaque domaine on demande avec qui il peut en parler, le type de relation avec l'interlocuteur, l'aide qui résulte du dialogue et s'il existe une réciprocité dans le soutien. Ce questionnaire nécessite la présence d'un cotateur pour orienter et vérifier que personne n'a été oublié dans l'établissement du réseau social. Trois scores peuvent

être calculés : la qualité, l'étendue du réseau et le degré de réciprocité. Les auteurs ont voulu tester le biais dû aux réponses « socialement désirables ». 19 étudiants ont rempli le SRS en 2 occasions et la deuxième fois, pour encourager le biais, il leur a été demandé de répondre en supposant des circonstances idéales. Les résultats ont été différents de manière significative ce qui a permis aux auteurs de conclure que les réponses sont indépendantes du facteur « socialement désirable ». On ne peut étendre cette conclusion que si l'on estime que 19 étudiants en psychologie représentent l'ensemble de la population. Ce questionnaire se remplit rapidement. Il donne des informations sur la qualité et la quantité des contacts sociaux, sur les membres du réseau sans préciser toutefois le type d'aide fournie. L'auteur a pu montrer que la qualité de l'aide a plus d'impact que le nombre de relations sociales.

The Social Support Questionnaire (SSQ) (Sarason, 1983)

Comme le précédent, le SSQ est un questionnaire qui mesure la qualité et la quantité du réseau social. Ce questionnaire de 27 items est composé de 2 parties ; le SSQN qui permet d'évaluer le nombre de personnes qui appartiennent au réseau et le SSQS qui permet de préciser le degré de satisfaction sur une échelle en 6 points. Un maximum de 9 personnes peut être listé en précisant leurs initiales et le lien de relation avec le sujet. Deux scores sont alors obtenus ; un score quantitatif et un score de satisfaction. Le coefficient alpha pour le SSQN est de 0,97 et pour le SSQS de 0,94. Les corrélations de test-retests sont respectivement de 0,9 et 0,83 après 4 semaines d'intervalle. Des corrélations négatives significatives ont été trouvées entre le SSQ et une échelle de dépression. Sur une population de 295 étudiants, une association significative a été montrée entre le nombre d'événements de vie positifs et le score au SSQ ; ceux qui avaient le plus d'événements de vie positifs avaient le plus haut score au SSQ. Ceux qui avaient le plus de soutien étaient les plus capables de faire face aux événements de la vie. Cet instrument a été élaboré avec soin et représente un bon instrument pour étudier le support social.

The Interview Schedule for Social Interaction (ISSI) (Henderson, 1980; Duncan-Jones, 1981)

Pour Henderson qui s'est inspiré de Bowlby, l'être humain est biologiquement programmé pour vivre de façon grégaire. Il y trouve 6 bénéfices : le sens de l'attachement et de la sécurité, l'intégration sociale, la possibilité d'altruisme, la réassurance sur sa propre valeur, le sentiment de solidité du lien et la possibilité d'aide si besoin. Sur ces principes a été conçu l'ISSI. Il s'agit d'un entretien de 45 minutes qui envisage le soutien social dans ses aspects quantitatifs et qualitatifs durant les 12 mois précédents. Sur chaque type particulier de relation sociale, il est demandé si elle existe, si la quantité de ce type de relation est suffisante et qui les fournit. Quatre scores sont obtenus : existence et adéquation de l'attachement — 8 et 12 items —, existence et adéquation de l'intégration sociale — 16 et 1 items. La consistance interne des 4 scores s'étend de 0,67 à 0,79. Le résultat du test-retest après 18 jours pour chaque score est de 0,71 à 0,76. L'ISSI discrimine entre des groupes *a priori* différents en matière d'ajustement social : les nouveaux arrivants dans une ville comparés aux résidents, les personnes divorcées ou séparées comparées aux personnes mariées. Une analyse factorielle montre que l'existence de l'attachement

et de l'intégration sociale est située sur deux axes relativement indépendants mais que les perceptions de l'adéquation de ces deux domaines sont étroitement corrélées. L'ISSI mesure plutôt le support social que le rôle social. Henderson a pu montrer que la qualité mieux que la quantité du réseau est meilleur prédicteur de la résistance aux troubles psychologiques. Les analyses empiriques de la structure de l'échelle n'ont pas coïncidé avec l'hypothèse théorique de départ. Les notions de possibilité d'altruisme, de réassurance sur sa propre valeur, de sentiment de solidité du lien ne se sont pas dégagées. Par contre une dimension plus générale d'intégration sociale a pu être formée. Cette reformulation montre bien que les instruments élaborés influent sur la définition et les limitations mêmes de l'objet d'étude.

The Social Interaction Schedule (SIS) (Henderson, 1978) et l'Interview Measure of Social Relationships (IMSR) (Brugha, 1987)

Antérieurement à l'ISSI, Henderson avait élaboré la SIS pour évaluer les interactions sociales dans des échantillons larges. Il s'agissait d'un entretien de 76 items qui collectait des informations sur la qualité et la quantité des contacts sociaux sur les 7 derniers jours précédant l'entretien. À partir de cet instrument un peu long à remplir (environ une heure), Brugha a élaboré l'IMSR qui repose sur les mêmes principes mais qui ne prend que 15 minutes à remplir.

Le Social Functioning Schedule (SFS) (Remington et Tyrer, 1979)

Le SFS est un entretien semi-structuré conçu pour évaluer les problèmes que rencontre un sujet dans les divers aspects du fonctionnement social. Cette échelle a été construite pour l'évaluation du traitement de sujets névrotiques lors de consultations externes. Le patient est interrogé sur les difficultés qu'il rencontre sans demander de précision sur ces difficultés. Le SFS inclut 12 sections : emploi, tâches ménagères, contribution à l'entretien de la maison, argent, entretien personnel, relation maritale, soins des enfants, relation patient-enfant, relation patient-parent, contacts sociaux, passe-temps et activités de loisirs. Les sections sans objet pour le patient peuvent être omises. L'examinateur est libre d'adapter ou d'ajouter des questions s'il le juge nécessaire pour obtenir suffisamment d'informations. Au total le SFS inclut 121 questions. En outre le cotateur dispose d'une échelle analogue visuelle pour évaluer le niveau de problème dans une section donnée. L'évaluation se rapporte aux quatre semaines écoulées et l'entretien prend 10 à 20 minutes.

The Social Dysfunction Rating Scale (SDRS) (Linn et al., 1979; Linn, 1988)

Le concept central de l'auteur est la dysfonction envisagée comme une absence de plaisir accompagnée par des attitudes de dévalorisation. La SDRS évalue les aspects négatifs de l'ajustement social et ce indépendamment du rôle social. Elle se présente sous la forme de 21 items qui sont chacun des questions semi-structurées combinant l'évaluation du cotateur à une auto-évaluation. Par exemple, le cotateur évalue les contacts sociaux amicaux puis on demande au sujet s'il a envie d'avoir plus d'amis.

Chaque item est mesuré sur une échelle de 6 points. Ils se regroupent en 3 classes : 4 items sont fonction de l'opinion du sujet, 6 items sont fonction de ses relations interpersonnelles et 11 items concernent son manque de succès et son insatisfaction dans les situations sociales. La fidélité intraclasses s'étale de 0,54 à 0,86. L'instrument a montré sa capacité à distinguer une population de schizophrènes *versus* des témoins, ce qui est intéressant mais pas vraiment essentiel pour ce type d'échelle. Cinq facteurs ont été identifiés par l'analyse factorielle : apathie/détachement, insatisfaction, hostilité, santé/finances, dépendance. La SDRS a été particulièrement étudiée pour l'évaluation des aspects du fonctionnement social chez la personne âgée. Cet instrument est court et envisage une vaste étendue de symptômes de dysfonctionnement.

The Structured and Scaled Interview to Assess Maladjustment (SSIAM) (Gurland, 1972)

Le dysfonctionnement social est ici aussi le concept de base. Il est défini par les conduites objectives et les sensations subjectives d'absence de satisfaction qui résultent des activités sociales. Conçue pour évaluer les résultats de la psychothérapie, la SSIAM fournit donc une évaluation clinique détaillée du dysfonctionnement social. Elle se présente sous la forme d'un entretien structuré de 60 items. 45 items sont groupés en 5 champs : travail, relations sociales, famille, mariage et sexualité. Les 15 items restants s'adressent au cotateur et lui permettent de donner son jugement sur le niveau de stress de l'environnement, son pronostic compte tenu de la volonté de changement du sujet. À l'intérieur de chaque domaine, outre le jugement du cotateur, trois types de dysfonctionnement sont mesurés; le comportement, les frictions et la détresse. Les 60 items sont cotés sur une ligne de 10 points dont un sur deux est défini. L'entretien dure 30 minutes. Les corrélations interclasses ont été calculées à partir de l'entretien de 15 patients par trois cotateurs. Elles vont de 0,78 à 0,97. Une analyse factorielle a mis en évidence que 21 items comptant pour 40 % de la variance sont répartis en 6 facteurs : isolation sociale, inadéquation au travail, friction avec la famille, dépendance avec la famille, insatisfaction sexuelle et friction en dehors de la famille. Pour 89 patients, un proche a été interviewé pour fournir une cotation indépendante sur les 6 facteurs identifiés par l'analyse factorielle. Les comparaisons ont montré des corrélations statistiquement significatives excepté sur le facteur dysfonctionnement sexuel. Les trois manifestations de dysfonctionnement, — comportement, friction et détresse — qui auraient dû apparaître au sein des 5 champs ne se sont pas distinguées lors de l'analyse fonctionnelle. La SSIAM est une échelle fréquemment utilisée et qui a été capable d'identifier des changements significatifs avant et après psychothérapie. Ses qualités reposent sur une clarté de méthodes et de concept.

The Social Adjustment Scale (SAS) (Weissman et al., 1971, Paykel et al., 1971, Weissman et al., 1976)

La SAS a été développée à partir de la SSIAM pour mesurer l'ajustement social des femmes déprimées. Il s'agit d'une échelle conçue pour mesurer les résultats d'une psychothérapie, où les intentions sont moins de soulager les symptômes que d'améliorer les relations interpersonnelles et les habiletés sociales. Il existe deux

versions de la SAS : un entretien et un questionnaire, le SAS-SR développé à partir du premier et dont les avantages sont la facilité de passation et l'absence de biais dû au cotateur. Les deux versions contiennent 42 questions couvrant les performances dans les rôles sociaux de 6 aires de fonctionnement : travail (qui se décompose en sous-chapitres), loisir social, relations avec la famille au sens large, rôle de conjoint, de parent et de membre de l'unité familiale. Dans chaque aire de fonctionnement, les questions ont trait à la quantité de contacts sociaux, aux frictions interrelationnelles, au niveau d'indépendance, aux sentiments personnels (timidité...) et au degré de satisfaction. Seules les deux semaines précédentes sont évaluées. Deux systèmes de cotations sont utilisés : un score moyen pour chaque section (travail, loisir) et un score global qui exclut toute référence au rôle social. Le SAS-SR prend 15 à 20 minutes à remplir en présence d'un assistant dont le rôle est de s'assurer que le document est compris et rempli en totalité. L'entretien dure une heure. Les scores ne sont pas significativement corrélés à l'âge, la classe sociale, le sexe, les antécédents de dépression. Une analyse factorielle a été faite sur la SAS-SR. Six facteurs ont été individualisés : performance au travail, friction interpersonnelle, communication inhibée, dépendance docile, attachement familial et rumination anxieuse. La comparaison entre les scores de la SAS et de la SAS-SR a montré que les patients se considéraient plus pauvres en relations sociales par rapport au jugement des cotateurs. La corrélation sur l'ajustement global était de 0,72 reflétant une bonne consistance. Le SAS-SR administré avant et après 4 semaines de traitement chez 76 déprimés voit ses scores modifiés dans le sens d'une amélioration. Des corrélations ont été effectuées pour différents types de population avec différentes échelles. Pour les patients schizophrènes, une SAS-II a été développée sous la forme d'un entretien semi-structuré de 56 items. Les questions s'adaptent au sujet qu'il travaille ou non. Mais s'il est trop malade pour avoir un rôle quelconque au travail, ou s'il commence un travail, son score sera bas et il apparaîtra comme déficitaire. Les études de validité et de fiabilité ont été réalisées sur différents types de population, ce qui n'est pas si fréquent. Les études métrologiques et leur résultat font que cet instrument est très largement utilisé en recherche.

Le DSM-IV (APA, 1994)

— L'axe IV du DSM-IV évalue les problèmes psychosociaux et environnementaux qui pourraient affecter le diagnostic, le traitement et le pronostic des troubles mentaux. Ces problèmes sont ainsi regroupés :

– Problèmes avec le groupe de support primaire (décès d'un membre de la famille, éclatement de la famille par séparation, divorce, ou désaffection...).

– Problèmes en rapport avec l'environnement social (décès ou perte d'un ami, support social inadéquat, habiter seul...).

– Problèmes d'éducation (illétrisme, problèmes scolaires...).

– Problèmes de travail (chômage, menace de perte d'emploi...).

– Problèmes de logement (absence de logement, logement inadéquat).

– Problèmes économiques (pauvreté extrême, finances inadéquates).

– Problèmes d'accès au service de santé (services de santé inadéquats...).

– Problèmes judiciaires (arrestation, incarcération...).

– Problèmes environnementaux divers (expositions à des catastrophes, guerres...).

Le clinicien devra tenir compte de problèmes psychosociaux et environnementaux de l'année précédant l'évaluation actuelle. Il peut néanmoins tenir compte de problèmes antérieurs s'ils contribuent de façon évidente aux troubles mentaux ou s'ils vont représenter un point central du traitement (par exemple les états de stress posttraumatiques).

— L'axe V du DSM-IV rend compte du jugement du clinicien sur le fonctionnement global du sujet. Cette information sert à la planification du traitement, à la mesure de son impact et à la prédiction du résultat. L'échelle utilisée est le *Global Assessment Functioning Scale* (GAF). Il s'agit d'un instrument en 100 points (et non plus en 90 points comme la GAF du DSM-III-R) qui combine les mesures des trois axes du fonctionnement adaptatif selon le DSM-IV : psychologique, social et occupationnel (à l'exclusion des limitations physiques et environnementales). L'évaluation devra s'effectuer sur la période en cours (le niveau de fonctionnement au moment de l'évaluation). Elle peut se faire cependant à différents moments (à l'entrée ou à la sortie d'une hospitalisation) ou sur une période passée (le meilleur niveau de fonctionnement sur au moins quelques mois durant l'année passée).

— La mesure combinée de l'adaptation sociale et de syndromes de l'axe I a été critiquée (Goldman *et al.*, 1992; Skodol *et al.*, 1988). Si l'approche multiaxiale doit permettre la quasi-indépendance de l'évaluation des différents domaines, l'inclusion explicite des syndromes dans l'axe V va à l'encontre de ce principe. De plus des études (Gordon *et al.*, 1988) ont montré que l'amélioration des symptômes et l'amélioration fonctionnelle n'évoluent pas de façon parallèle mais indépendante. Les auteurs du DSM-IV proposent l'évaluation de l'incapacité sociale et occupationnelle et le suivi des progrès de la réadaptation indépendamment de la sévérité des symptômes psychologiques. L'échelle proposée est la *Social and Occupational Functioning Assessment Scale* (SOFAS). Il s'agit d'une échelle en 100 points dérivée de la *Global Assessment Scale* (Endicott *et al.*, 1976).

— Le DSM-IV propose un axe additionnel qui permet au clinicien d'indiquer son jugement global sur le fonctionnement de la famille ou d'une autre relation en cours le long d'un continuum hypothétique s'étendant d'un fonctionnement compétent, relationnel optimal à une relation dysfonctionnelle. L'échelle utilisée est la *Global Assessment of Relational Functioning Scale* (GARF). La GARF permet de coter le degré auquel une famille ou une autre unité relationnelle actuellement en cours répond au besoin affectif ou instrumental de ses membres dans les domaines suivants : résolution de conflits, organisation, climat émotionnel. La GARF devrait être utilisée pour coter le fonctionnement durant la période actuelle.

CONCLUSION

L'adaptation sociale n'est pas un concept simple, et ceci justifie la multiplicité des instruments de mesure. mais le travail du chercheur ne s'arrête pas à leur élaboration et seules des études métrologiques — nombreuses et renouvelées par des auteurs différents, sur des échantillons de taille raisonnable, sur des populations variées... —, permettront leur utilisation raisonnable et appropriée.

BIBLIOGRAPHIE

BRUGHA T. S., STURT E., MCCARTHY B. *et al.* — The Interview Measure of Social Relationships : the Description and Evaluation of a Survey Instrument for Assessing Personal Resources. *Soc. Psychiatry, 22*, 123-128, 1987.

CLARE A. W., CAIRNS V. E. — Design, Development and Use of a Standardized Interview to Assess Social Maladjustment and Dysfunction in Community Studies. *Psychol. Med., 8*, 589-604, 1978.

DUNCAN-JONES P. — The Structure of Social Relationship : Analysis of a Survey Instrument, Part I. *Soc. Psychiatry, 16*, 55-61, 1981.

DUNCAN-JONES P. — The Structure of Social Relationship : Analysis of a Survey Instrument, Part II. *Soc. Psychiatry, 16*, 143-149, 463-466, 1981.

ENDICOTT J., SPITZER R. L., FLEISS J. L. *et al.* — The Global Assessment Scale : a Procedure for Measuring Overall Severity of Psychiatric Disturbance. *Arch. Gen. Psychiatry, 33*, 766-771, 1976.

GOLDMAN H. H., SKODOL A. E., LAVE T. R. — Revising Axis V for DSM-IV : a Review of Measures of Social Functioning. *Am. J. Psychiatry, 149*, 1148-1156, 1992.

GORDON R. E., PLUTZKY M., GORDON K. K. *et al.* — Using the Axis V scale to Evaluate Therapeutic Outcome of Psychiatric Treatment. *Can. J. Psychiatry, 33*, 194-196, 1988.

GURLAND B. J., YORSTONE N. J., STONE A. R. *et al.* — *Structured and Scaled Interview to Assess Maladjustment* (SSIAM). I. Description, Rationale and Development. *Arch. Gen. Psychiatry, 27*, 259-263, 1972.

GURLAND B. J., YORSTONE N. J., STONE A. R. *et al.* — *Structured and Scaled Interview to Assess Maladjustment* (SSIAM). I. Springer, New York, 1974.

HENDERSON S., DUNCAN-JONES P., BYRNE D. G. *et al.* — Measuring Social Relationship : the Interview Schedule for Social Interaction. *Psychol. Med., 10*, 723-734, 1980.

HENDERSON S., DUNCAN-JONES P., BYRNE D. G. *et al.* — Social Bonds in the Epidemiology of Neurosis; a Preliminary Communication. *Br. J. Psychiatry, 132*, 463-466, 1978.

KATZ M. M., LYERLY S. B. — Methods of Measuring Adjustment and Social Behavior in the Community. *Psychol. Rep., 13*, 503-535, 1963.

LEWIS J. — Health as a Social Concept, *Brit. J. Sociol., 4*, 109-124, 1953.

LINN M. W., SCULTHORPE W. B., EVJE M. *et al.* — A Social Dysfunction Rating Scale. *J. Psychiatr. Research, 6*, 299-306, 1969.

LINN M. W. — A Critical Review of Scales Used to Evaluate Social and Interpersonal Adjustment in the Community. *Psychopharmacol. Bull., 24*, 615-621, 1988.

MCFARLANE A. H., NEALE K. A., NORMAN G. R. *et al.* — Methodological Issues in Developing a Scale to Measure Social Support. *Schizophr. Bull., 7*, 90-100, 1981.

PAYKEL E. S., WEISSMAN M. M., PRUSOFF B. A. *et al.* — Dimensions of Social Adjustment in Depressed Women. *J. Nerv. Dis., 152*, 158-172, 1971.

REMINGTON M., TYRER P. — The Social Functioning Schedule – a Brief Semi-Structured Interview. *Soc. Psychiatry, 14*, 151-157, 1979.

SARASON I. G., LEVINE H. M., BAHSAM R. B. *et al.* — Assessing Social Support : the Social Support Questionnaire. *J. Pers. Soc. Psychol., 44*, 127-139, 1983.

SKODOL A. E., LINK B. G., SCHROUT P. E. *et al.* — The Revision of Axis V in DSM-III-R : should Symptoms have been Included. *Am. J. Psychiatry, 145*, 825-829, 1988.

WEISSMAN M. M., BOTHWELL S. — Assessment of Social Adjustment by Patient Self-Report. *Arch. Gen. Psychiatry, 33*, 1111-1115, 1971.

WEISSMAN M. M., KLERMAN G. L., PAYKEL E. S. — Clinical Evaluation of Hostility in Depression. *Am. J. Psychiatry, 128*, 261-266, 1971.

24

LE «*EATING DISORDER INVENTORY*» (EDI)

par S. CRIQUILLION-DOUBLET, S. DIVAC, R. DARDENNES et J. D. GUELFI

INTRODUCTION

À l'heure actuelle, de nombreux instruments d'évaluation des troubles du comportement alimentaire ayant des caractéristiques métrologiques satisfaisantes sont rapportés dans la littérature. La plupart d'entre eux provenant de pays anglophones, leur utilisation en France demeure limitée en l'absence d'un nombre suffisant d'études de validation. De plus, leur utilisation ne remplace en aucun cas la démarche diagnostique entreprise au cours des entretiens, que ceux-ci soient ou non standardisés, mais elle permet une quantification des données recueillies et l'évaluation de l'intensité des troubles au cours des différentes phases évolutives.

Un des rares instruments d'évaluation des troubles du comportement alimentaire validés sur une population française est le «*Eating Attitudes Test*» (EAT) conçu par Garner et Garfinkel en 1979 (Carrot *et al.*, 1986). Destiné au dépistage de l'anorexie mentale, il est surtout utilisé pour l'évaluation de l'intensité des symptômes. L'expérience clinique a cependant démontré l'insuffisance de ce questionnaire concernant, en particulier, l'évaluation des aspects psychologiques des conduites alimentaires.

Nous présentons ici un inventaire élaboré en 1983 par Garner et ses collaborateurs, le *Eating Disorder Inventory* (EDI), ainsi que les premiers résultats d'une étude française réalisée à la Clinique des maladies mentales et de l'encéphale (Pr B. Samuel-Lajeunesse).

PRÉSENTATION DE L'ÉCHELLE

Historique

Le *Eating Disorder Inventory* (EDI) a été construit par Garner, Olmsted et Polivy afin de disposer d'un instrument permettant d'étudier les caractéristiques cognitives et comportementales des sujets présentant des troubles du comportement alimentaire de type anorexie mentale ou boulimie.

L'EDI a déjà été traduit en plusieurs langues : le suédois, le hollandais, l'allemand, l'espagnol, le chinois, le français, et les premières études transculturelles commencent à apparaître. Dans sa première version, l'EDI traduit en français par J. D. Guelfi et S. Criquillion-Doublet est utilisé à la Clinique des maladies mentales

et de l'encéphale (CMME) depuis mai 1990. Composé de 64 items, cet inventaire a récemment été étendu à 91 items (EDI-2, Garner, 1991). Ce questionnaire permet de couvrir les aspects cognitifs et comportementaux de différents sous-groupes de patients : anorexiques restrictifs et boulimiques, patients boulimiques à poids normal, obèses, mais aussi cas sub-syndromiques, infracliniques, et sujets « normaux ».

Cet EDI a été extrait d'un groupe de 146 items formulés par des praticiens qui connaissent les données de la littérature sur l'anorexie mentale et la boulimie et qui étaient familiers de la clinique et de la prise en charge de tels patients. Ces items ont été élaborés pour mesurer 11 domaines clés de la pathologie ; 8 d'entre eux ont satisfait aux critères de fiabilité et de validité sur l'échantillon de population étudié. Ce questionnaire consiste en une auto-évaluation de type papier-crayon à partir de 64 propositions pour lesquelles le sujet doit répondre en cochant l'un des six degrés suivants : *« toujours »*, *« habituellement »*, *« souvent »*, *« quelquefois »*, *« rarement »*, ou *« jamais »*. Une passation directe sur ordinateur est de plus en plus souvent utilisée pour un dépouillement direct. Les scores vont de 0 à 3, la note 3 représentant l'extrême de l'échelle, la réponse *« toujours »* ou *« jamais »* selon la nature de la question allant soit dans le sens du symptôme, soit, à l'inverse, dans le sens de l'absence de symptôme.

La note de 2 est attribuée aux réponses *« habituellement »* ou *« rarement »*, et le score 1 aux deux dernières propositions, *« souvent »* ou *« quelquefois »*, selon le même principe. Les 3 dernières possibilités de réponse sont toujours cotées 0, en allant de droite à gauche ou *vice versa*, en fonction de la nature *« symptomatique »* ou *« non symptomatique »* des questions. Certaines données cliniques et biographiques complètent la passation telles que l'âge, le statut marital, le poids actuel, le poids minimal, maximal et leur durée de maintien, le poids idéal selon le patient, l'âge de début des troubles, et le statut professionnel du patient et de ses parents.

DESCRIPTION DES SOUS-ÉCHELLES

Désir intense de minceur

Cette sous-échelle est souvent considérée comme une des principales dimensions des troubles des conduites alimentaires. Témoignent du désir de minceur ou de la crainte de la corpulence les régimes alimentaires répétés et les préoccupations persistantes centrées sur le poids.

Boulimie

Les préoccupations sur la nourriture conduisent parfois à des épisodes de suralimentation particuliers, suivis dans certains cas par des vomissements. La présence de cette boulimie permet de distinguer ces patients de sujets uniquement restrictifs ; elle est généralement considérée comme un facteur de mauvais pronostic. Les accès boulimiques sont également décrits chez des jeunes filles ou des jeunes femmes n'ayant aucun antécédent d'anorexie mentale.

Insatisfaction corporelle

Elle correspond à la croyance que certaines parties du corps, changeant de forme au moment de la puberté, deviennent trop importantes, comme les hanches, les cuisses, les fesses. Cette échelle est étroitement liée aux troubles de l'image du corps, longtemps considérés comme une des principales composantes de ces troubles.

Inefficacité

Les sentiments d'insécurité, de dévalorisation, évalués par cette échelle, sont étroitement liés à une mauvaise image de soi et à des sentiments d'inadéquation. Certains auteurs, comme Hilde Bruch ou Mara Selvini-Palazzoli, considèrent ce sentiment d'inefficacité personnelle comme une perturbation fondamentale de l'anorexie mentale.

Perfectionnisme

Cette échelle évalue la tendance à imposer à soi-même et, quelquefois, aux autres des exigences (« standards ») élevées ou excessives. Selon Garner *et al.* (1983), le perfectionnisme fait partie du style de pensée dichotomique.

Méfiance dans des relations interpersonnelles

Le sujet se méfie des relations étroites avec d'autres personnes de peur d'exprimer librement ses émotions. Cette méfiance jouerait un rôle important dans le développement de l'anorexie mentale.

Conscience intéroceptive

Cette sous-échelle se rapporte au manque de confiance dans la reconnaissance de ses émotions ou de ses sensations corporelles, telles que la satiété ou la faim.

Peur de la maturité

Cette sous-échelle évalue la tendance du sujet à toujours rechercher un sentiment de sécurité de type infantile par crainte des exigences de l'âge adulte. Pour Arthur Crisp (1965, 1981), la psychopathologie centrale de l'anorexie mentale se situe justement autour de cet évitement de la maturité biologique et psychologique. L'apparente satisfaction qu'apporte le maintien des formes corporelles d'un enfant, avec perte des formes féminines et aménorrhée, en est un exemple.

Les 3 nouvelles échelles incluses dans l'EDI-2 concernent l'ascétisme, le contrôle des impulsions et le sentiment d'insécurité sur le plan des relations sociales.

Ascétisme

Cette sous-échelle recouvre la valorisation de l'autocontrôle et de la discipline concernant les besoins corporels jusqu'à la privation de tout plaisir.

Contrôle des impulsions

Le score de cette sous-échelle reflète la tendance à agir d'une manière impulsive dans les relations interpersonnelles, à exprimer de l'hostilité et à manquer d'autocontrôle, ce qui est habituellement considéré comme un signe de mauvais pronostic.

Insécurité sur le plan social

Cette sous-échelle estime les croyances selon lesquelles les relations sociales ne sont guère fiables et sont la source de déceptions; ces croyances sont souvent à l'origine d'un évitement des relations sociales et nécessitent parfois des prises en charges spécifiques.

ÉTUDES DE VALIDATION DE L'EDI

L'étude initiale (Garner *et al.*, 1983) a été effectuée en deux temps: 113 patients anorexiques, dont 65 restrictifs et 48 boulimiques ont été comparés à 577 sujets contrôles. L'âge et la durée de maladie des patients anorexiques (selon les critères de Feighner *et al.*, 1972) ne différaient pas selon le sous-type (restrictif pur ou boulimique), et le poids était plus élevé chez les patients boulimiques. Les 577 sujets féminins contrôles ont été recrutés parmi des étudiants en psychologie.

Les coefficients de cohérence interne — alpha de Cronbach — étaient satisfaisants pour le groupe des anorexiques, avec des valeurs supérieures à .80 (moindres mais néanmoins supérieures à .60 chez les sujets contrôles).

Cette étude initiale a été complétée par les données obtenues pour d'autres troubles des conduites alimentaires et pour différentes populations : 195 patients boulimiques à poids normal, 44 patients actuellement obèses, 52 patients antérieurement obèses, 17 patients ayant des antécédents d'anorexie et 166 hommes étudiants formant un groupe contrôle. Les personnes ayant des accès boulimiques ainsi que les obèses ont eu des scores plus élevés concernant l'insatisfaction corporelle et la boulimie que les patients anorexiques.

La valeur moyenne et l'écart type des scores des patients anorexiques, des anciens patients anorexiques, et des sujets-contrôle (féminin et masculin) de cette étude sont présentés sur le tableau 1.

La validité de chaque sous-échelle a également été testée en étudiant les corrélations entre les évaluations par EDI et les évaluations des cliniciens. Les coefficients de corrélation entre les scores des sous-échelles et le jugement des cliniciens étaient tous significatifs, variant de 0,43 à 0,68. De plus, une analyse discriminante a montré que l'EDI permettait de distinguer les sujets anorexiques et les sujets contrôle avec des pourcentages de sujets correctement classés de 88 % à 92 % selon la sous-échelle.

Tableau 1. *Notes moyennes (± erreur standard) des groupes : anorexie mentale, anorexie mentale guérie, contrôles femmes et contrôles hommes.*
In EDI : Garner D. M., Olmsted M. P., Polivy J. (1983)

	Anorexia Nervosa (N = 113)*	*Recovered Anorexia Nervosa (N = 17)*	*Female Comparison (N = 577)*	*Male Comparison (N = 166)*
Désir de minceur	15,4 (± 0,50)	3,6 (± 1,3)	5,0 (± 0,22)	1,6 (± 0,24)
Boulimie	R = 2,7 (± 0,55) B = 10,8 (± 0,69)	0,3 (± 0,14)	2,0 (± 0,14)	1,0 (± 0,14)
Insatisfaction corporelle	R = 14,2 (± 1,0) B = 10,8 (± 1,0)	6,3 (± 1,5)	10,2 (± 0,32)	3,9 (± 0,39)
Inefficacité	14,4 (± 0,75)	3,1 (± 0,95)	2,0 (± 0,15)	1,6 (± 0,24)
Perfectionnisme	10,0 (± 0,47)	6,5 (± 0,88)	5,2 (± 0,16)	6,2 (± 0,30)
Méfiance interpersonnelle	7,7 (± 0,49)	1,9 (± 0,56)	2,2 (± 0,12)	3,1 (± 0,24)
Conscience intéroceptive	12,5 (± 1,1)	2,1 (± 0,86)	2,9 (± 0,47)	1,4 (± 0,39)
Peur de la maturité	6,0 (± 0,99)	1,1 (± 0,46)	2,5 (± 0,33)	2,7 (± 0,44)

* R = Restricters (N = 48) ; B = Bulimics (N = 65) ; when bulimic and restricter subgroups do not differ significantly on subscale scores, only total group means are reported.
+ N = 35 for AN, N = 13 for Recovered AN, N = 82 for FC, N = 68 for MC.

Structure factorielle

L'étude de Welch, Hall et Norring (1990), menée sur 271 patientes ayant des troubles des conduites alimentaires, en Suède et en Nouvelle-Zélande, a confirmé l'existence de 8 facteurs sur les onze dimensions initialement décrites par Garner.

Plusieurs études de ce type ont été menées en Suède et en Nouvelle-Zélande (Norring et Sohlberg, 1988, Welch *et al.*, 1990) avec des résultats tout à fait comparables pour 6 sous-échelles qui s'avéreraient donc les plus stables, les sous-échelles « Perfectionnisme » et « Méfiance interpersonnelle » faisant exception.

L'existence de ces 8 facteurs a également été confirmée dans une étude de Williams *et al.* (1986) avec un groupe contrôle.

Stabilité de l'EDI

Peu de données sont actuellement disponibles pour juger de la stabilité de l'inventaire selon les échantillons étudiés. Les scores des femmes paraissent plus élevés sur les sous-échelles concernant le corps, l'apparence physique, la conscience intéroceptive et la boulimie. Quelques études concernant les différences entre les étudiantes canadiennes et allemandes, par exemple, (Steinhausen, 1984), ont montré que les étudiantes canadiennes avaient des scores plus élevés sur la plupart des échelles que les allemandes, à l'exception de la sous-échelle : peur de la maturité.

INTÉRÊT ET LIMITES DE L'EDI

L'instrument paraît suffisamment sensible au changement obtenu sous traitement, et plusieurs études sur la valeur prédictive sont déjà publiées. L'adjonction de 3 sous-échelles dans la nouvelle version de l'EDI est encore trop récente pour que l'on puisse se prononcer sur leur valeur véritable. Enfin, comme toute échelle d'auto-évaluation, le questionnaire EDI peut malheureusement faire l'objet de réponses biaisées du fait du déni souvent observé chez ces patients, ce qui souligne l'intérêt de corréler les données de l'examen clinique à celles obtenues lors de la passation d'instruments d'évaluation.

VALIDATION DE LA VERSION FRANÇAISE DE L'EDI

Les résultats présentés constituent la première étape des travaux de validation de la version française du questionnaire EDI.

Méthode

Cette étude a été réalisée chez 193 sujets ayant des troubles du comportement alimentaire, hospitalisés ou suivis à la Clinique des maladies mentales et de l'encéphale depuis 1991. Il s'agit essentiellement de femmes (187 sur 193) réparties en trois groupes principaux selon les critères diagnostiques du DSM-III-R :

– 62 sujets anorexiques restrictifs,

– 67 sujets boulimiques à poids normal,

– 56 sujets répondant à la fois aux critères diagnostiques de l'anorexie et de la boulimie (anorexiques-boulimiques).

Deux autres groupes diagnostiques — «*binge eating disorders*» et «hyperphages et autres troubles des conduites alimentaires» — comprennent très peu de sujets (quatre chacun) et ne seront pas analysés du fait du faible effectif. La passation de l'EDI a été effectuée à l'issue de deux premières consultations tant pour les sujets suivis en ambulatoire que pour les patients hospitalisés (consultations de pré-admission dans ce dernier cas). Cette passation fait partie intégrante d'une évaluation clinique globale comprenant également un recueil de données anamnestiques, divers autres questionnaires d'auto-évaluation (inventaire de dépression de Beck, questionnaires alimentaires, *Eating Attitudes Test* (EAT), *Symptom Checklist* (SCL-90-R) ainsi que plusieurs échelles d'hétéro-évaluation comme les échelles de dépression et d'anxiété de Hamilton et l'échelle de Morgan et Russell.

Caractéristiques de l'échantillon (tableau 2)

L'âge moyen au moment de la passation de l'EDI des trois groupes de sujets — anorexiques, boulimiques et anorexiques-boulimiques — est comparable et se situe autour de 26 ans.

Tableau 2. *Caractéristiques de l'échantillon*

	Anorexiques *n = 62*		*Boulimiques* *n = 67*		*Anorexiques-boulimiques* *n = 56*		*F*	*p* (*d.d.l = 2, 182)*
Âge (écart-type)	26,1	(7,3)	26,2	(6,5)	27,8	(6,4)	1,23	0,295
Âge de début des troubles	18,5	(5,6)	15,2	(3,9)	17,4	(3,8)	8,78	$0{,}2 \cdot 10^{-3}$
BMI actuel	14,6	(2,9)	20,0	(4,6)	15,1	(2,7)	46,55	$0{,}1 \cdot 10^{-14}$
BMI antérieur minimal	12,8	(2,2)	16,7	(3,3)	13,4	(1,9)	41,92	$0{,}1 \cdot 10^{-14}$
BMI antérieur maximal	21,3	(3,9)	24,0	(3,8)	21,9	(3,0)	9,75	$0{,}1 \cdot 10^{-3}$
BMI idéal	17,7	(1,4)	18,5	(2,1)	17,8	(1,5)	4,37	0,014
Taille	162,9	(7,1)	165,1	(7,0)	163,3	(7,1)	1,76	0,176

En revanche, l'âge moyen du début des troubles alimentaires n'est pas le même selon le groupe diagnostique; les boulimiques ont un âge de début plus précoce que celui des anorexiques, qu'elles soient restrictives pures ou anorexiques-boulimiques (15,2 ans en moyenne contre 18,5 et 17,4 ans).

Le BMI moyen (*body mass index* : rapport du poids (en kg) sur le carré de la taille (en m), normalement compris entre 19 et 24) est voisin pour les anorexiques restrictifs (14,6) et les anorexiques-boulimiques (15,1), tandis que le groupe des boulimiques à poids normal a un BMI significativement plus élevé (20,0). Il en est de même lorsqu'on examine le BMI antérieur minimal et le BMI antérieur maximal dans les trois groupes.

L'estimation du BMI idéal est inférieure aux valeurs normales dans les trois groupes et les deux groupes d'anorexiques l'estiment moins élevé que le groupe des boulimiques (proche de 17,8 contre 18,5).

Analyses statistiques

Le but principal a été de vérifier que la traduction de l'EDI et son application sur une population clinique française conservent la structure en 8 facteurs établie par les études de validation initiale de cet outil. Pour cela, une procédure identique à l'analyse de Welch *et al.* (1990) a été suivie :

1) *Calcul des coefficients de cohérence interne de chaque sous-échelle.* Si l'on se fonde sur les études antérieures effectuées sur des populations cliniques (par exemple : Garner *et al.*, 1983; Welch *et al.*, 1990), les valeurs attendues des coefficients alpha de Cronbach devraient être supérieures à 0,80.

2) *Analyse factorielle.* Une analyse factorielle exploratoire avec rotation varimax a été effectuée afin de savoir si les 8 sous-échelles proposées par leurs auteurs étaient identifiables dans la structure factorielle ainsi établie. Les seuils ont été fixés comme suit : les facteurs retenus devaient avoir une valeur propre (*eigenvalue*) supérieure à 1 (critère de Kayser) et les coefficients de saturation des items une valeur d'au moins 0,4.

Enfin, il s'agit de savoir si l'EDI permet de différencier les troubles des conduites alimentaires à travers les réponses des patients. Pour cela, les scores des 3 principaux

groupes de troubles ont été comparés au moyen d'analyses de variance suivies de comparaisons 2 à 2 *post hoc* (test de Scheffé) avec un risque alpha fixé à 5 %.

Les calculs ont été réalisés avec le logiciel Systat 5.0.

Résultats

Cohérence interne des sous-échelles de l'EDI (tableau 3)

La cohérence interne des 8 échelles est satisfaisante puisque toutes ont un coefficient alpha d'au moins 0,80. Les meilleurs résultats sont obtenus pour les échelles boulimie (alpha = 0,93), désir de minceur, insatisfaction corporelle et inefficacité (0,89), tandis que l'échelle perfectionnisme obtient le coefficient le plus bas (0,80).

Tableau 3. *Cohérence interne des sous-échelles*

Sous-échelle	*α de Cronbach*
Désir de minceur	0,89
Boulimie	0,93
Insatisfaction corporelle	0,89
Inefficacité	0,89
Perfectionnisme	0,80
Méfiance interpersonnelle	0,84
Conscience intéroceptive	0,84
Peur de la maturité	0,85

Structure factorielle de l'EDI (tableaux 4 et 5)

Il n'existe pas de facteur général et huit facteurs ont une valeur propre supérieure à 1, expliquant 49 % de la variance totale. Chacun de ces facteurs comprend une majorité d'items issus d'une seule sous-échelle. 54 items sur 64 sont correctement affectés, 4 sont classés dans un autre facteur (items 64, 54, 26, 2), 1 item est présent sur 2 facteurs (34), et 6 items ne sont affectés à aucun facteur (47, 40, 18, 56, 20, 6).

Ainsi, le premier facteur comprend tous les items de la sous-échelle boulimie et deux items supplémentaires; l'un est issu de l'échelle conscience intéroceptive (*« Quand je suis contrarié(e), j'ai peur de commencer à manger »*) et l'autre de l'échelle insatisfaction corporelle (*« Je trouve que mon estomac est trop gros »*). Ce facteur explique à lui seul 21,5 % de la variance totale.

Le deuxième facteur regroupe 6 items sur les 7 de la sous-échelle méfiance interpersonnelle et un item issu de la sous-échelle conscience intéroceptive (*« J'arrive à identifier clairement les émotions que je ressens »*). Un item (*« J'ai du mal à exprimer mes émotions aux autres »*) est également affecté au facteur 8 (*cf. infra*).

Les cinq facteurs suivants regroupent uniquement des items appartenant à une seule sous-échelle, respectivement « insatisfaction corporelle », peur de la maturité », « désir de minceur », « perfectionnisme » et « inefficacité ».

Le dernier facteur contient une majorité d'items de la sous-échelle conscience intéroceptive (6/10) et deux items issus de la sous-échelle méfiance interpersonnelle : *« J'ai besoin de garder une certaine distance avec les gens (je me sens mal à l'aise si quelqu'un essaie de trop se rapprocher de moi »)* et *« J'ai du mal à exprimer mes émotions aux autres »*.

Enfin, 6 items ne peuvent être classés (coefficients de saturation inférieurs à 0,4), trois items de la sous-échelle inefficacité, deux de la sous-échelle conscience intéroceptive et un de la sous-échelle peur de la maturité.

Tableau 4. *Analyse factorielle de l'EDI vf, 1994*
(valeurs propres et % de variance expliquée par les 8 facteurs)

Facteur	*Valeur propre (eigenvalue)*	*% de variance totale expliquée*
1	13,74	21.46
2	5,20	8.11
3	3,30	5.14
4	2,43	3.80
5	2,00	3.13
6	1,90	2.97
7	1,31	2.05
8	1,24	1.93

Tableau 5. *Analyse factorielle de l'EDI vf, 1994*
(coefficients de saturation (> 0,4)

Item	*Sous-échelle*	*F1*	*F2*	*F3*	*F4*	*F5*	*F6*	*F7*	*F8*
It 5	Boulimie	0,90							
It 38	Boulimie	0,84							
It 28	Boulimie	0,79							
It 46	Boulimie	0,72							
It 61	Boulimie	0,69							
It 4	Boulimie	0,68							
It 53	Boulimie	0,62							
It 64	Intéroc.	0,54							
It 2	Corps	0,45							
It 23	Méfiance		0,71						
It 57	Méfiance		0,61						
It 15	Méfiance		0,57						
It 30	Méfiance		0,56						
It 17	Méfiance		0,51						
It 34	Méfiance		0,43						0,41
It 26	Intéroc.		0,49						
It 55	Corps			0,76					
It 62	Corps			0,71					
It 59	Corps			0,68					
It 45	Corps			0,68					
It 19	Corps			0,64					
It 9	Corps			0,57					
It 31	Corps			0,57					
It 12	Corps			0,46					

Tableau 5. *Analyse factorielle de l'EDI vf, 1994*
(coefficients de saturation (> 0,4)

Item	*Sous-échelle*	*F1*	*F2*	*F3*	*F4*	*F5*	*F6*	*F7*	*F8*
It 22	Maturité				0,77				
It 39	Maturité				0,77				
It 58	Maturité				0,62				
It 14	Maturité				0,59				
It 48	Maturité				0,52				
It 3	Maturité				0,52				
It 35	Maturité				0,48				
It 16	Minceur					0,81			
It 49	Minceur					0,72			
It 32	Minceur					0,66			
It 25	Minceur					0,54			
It 1	Minceur					0,51			
It 11	Minceur					0,51			
It 7	Minceur					0,48			
It 43	Perfect.						0,62		
It 29	Perfect.						0,60		
It 52	Perfect.						0,52		
It 13	Perfect.						0,48		
It 36	Perfect.						0,48		
It 63	Perfect.						0,40		
It 50	Innef.							0,73	
It 41	Innef.							0,70	
It 27	Innef.							0,63	
It 42	Innef.							0,61	
It 10	Innef.							0,55	
It 37	Innef.							0,54	
It 24	Innef.							0,53	
It 21	Intéroc.								0,58
It 44	Intéroc.								0,57
It 60	Intéroc.								0,54
It 51	Intéroc.								0,53
It 8	Intéroc.								0,49
It 33	Intéroc.								0,47
It 54	Méfiance								0,44
It 47	Intéroc.								
It 40	Intéroc.								
It 18	Innef.								
It 56	Innef.								
It 20	Innef.								
It 6	Maturité								

Dans l'ensemble, les 8 sous-échelles de l'EDI peuvent être clairement identifiées dans le profil des coefficients de saturation des items sur les 8 facteurs établis par l'analyse statistique.

Comparaison des scores entre les différents types de troubles (tableau 6)

Tableau 6. *Scores des 3 groupes diagnostiques à l'EDI*

	Anorexiques *n = 62*		*Boulimiques* *n = 67*		*Anorexiques-boulimiques* *n = 56*		*F*	*p* (*d.d.l = 2, 182*)
Score total (e.t.)	63,1	(28,7)	88,2	(34,8)	78,3	(29,4)	10,44	$0{,}5 \cdot 10^{-4}$
Désir de minceur	9,6	(6,5)	14,2	(5,9)	11,4	(6,7)	8,45	$0{,}3 \cdot 10^{-3}$
Boulimie	2,0	(3,7)	12,1	(6,1)	8,4	(6,8)	52,00	$0{,}1 \cdot 10^{-14}$
Insatisfaction corporelle	12,1	(6,2)	16,1	(7,9)	12,3	(6,3)	6,85	0,001
Inefficacité	10,0	(6,9)	12,3	(7,6)	13,4	(7,4)	3,45	0,034
Perfectionnisme	6,5	(4,1)	7,9	(4,8)	7,6	(4,4)	1,67	0,191
Méfiance interpersonnelle	7,8	(5,0)	6,0	(4,3)	6,2	(4,3)	3,02	0,051
Conscience intéroceptive	9,6	(6,6)	13,1	(6,8)	11,4	(6,4)	4,67	0,011
Peur de la maturité	5,5	(5,1)	6,5	(5,0)	7,8	(5,6)	2,75	0,066

L'analyse du score global de l'EDI, rarement discutée dans les travaux précédents sur ce sujet, montre une différence importante et significative entre les trois groupes. Les anorexiques restrictives ont un score nettement et significativement plus bas que les groupes boulimiques et anorexiques-boulimiques, ces derniers ne se distinguant pas entre eux. Les scores obtenus aux différentes sous-échelles montrent des différences inter-groupes significatives pour 5 sous-échelles sur les 8 de l'EDI. Les sujets boulimiques se distinguent nettement des anorexiques restrictives par leurs scores significativement plus élevés aux sous-échelles *Désir de minceur*, *Insatisfaction corporelle*, et *Conscience intéroceptive* (test de Scheffe, $p < 0{,}05$). Les anorexiques-boulimiques donnent des réponses qui les rapprochent de l'un ou l'autre des groupes de comparaison selon la sous-échelle : leur score de *Désir de minceur* les distinguent des anorexiques restrictives (et ne diffère pas de celui des boulimiques) tandis que c'est l'inverse pour la sous-échelle *Insatisfaction corporelle*. Ces deux groupes d'anorexiques diffèrent quant à leur sentiment d'inefficacité, les anorexiques-boulimiques ayant des scores plus élevés que les anorexiques-restrictives (le score intermédiaire des boulimiques ne présente pas de différence significative avec ceux de ces deux groupes). Les trois groupes se distinguent nettement les uns des autres sur la sous-échelle boulimie dans le sens attendu, à savoir des scores élevés pour les boulimiques, bas pour les anorexiques-restrictives et intermédiaires pour les anorexiques-boulimiques.

En revanche, les trois groupes ne se distinguent pas sur les sous-échelles *Perfectionnisme*, *Méfiance interpersonnelle* et *Peur de la maturité*.

DISCUSSION

L'évaluation de la consistance interne des sous-échelles de l'EDI établies initialement à partir de constatations empiriques effectuées par des cliniciens expérimentés a montré la fiabilité de l'instrument comme en témoignent d'autres études de référence (Welch *et al.*, 1988 et 1990; Hurley, 1990).

L'analyse factorielle réalisée sur les 187 patients de notre étude plaide aussi en faveur de l'existence d'une structure à 8 facteurs. Ces facteurs correspondent aux huit sous-échelles cliniques précitées.

Cette structure factorielle déjà mise en évidence dans d'autres travaux, chez des sujets présentant des troubles alimentaires, n'a pas été retrouvée chez des sujets témoins. Dans cette population, l'analyse des données de l'EDI donne une solution à trois facteurs : le désir de minceur, l'estime de soi, l'insatisfaction corporelle ne veulent pas dire la même chose pour tous.

Que peuvent en effet représenter la boulimie, le perfectionnisme pathologique et la conscience intéroceptive dans un groupe « contrôle » ?

Cependant, l'existence de différences significatives recueillies entre les différents sous-groupes d'anorexiques et de boulimiques doit nous permettre de mieux appréhender les aspects cognitifs spécifiques associés à ces troubles et les traits psychopathologiques caractéristiques qui s'y rattachent.

Des études ultérieures sont nécessaires pour répondre aux diverses questions que nous nous posons à l'issue de cette étape préliminaire de la validation de la traduction française de l'EDI. La sensibilité de l'instrument au changement est-elle la même selon les sous-groupes de patients? Pourra-t-on mettre en évidence des éléments de prédictivité des troubles, notamment pour la survenue de la boulimie? Comment tenir compte du fait que les notes obtenues à l'EDI varient en fonction de l'âge, du sexe et du degré de déni des troubles éventuels?

Ces réflexions marquent sans doute les limites des questionnaires de ce type susceptibles d'aider à l'évaluation du changement sous traitement et à la différenciation de groupes de patients atteints de troubles distincts mais qui ne doivent pas être employés dans un but diagnostique.

BIBLIOGRAPHIE

American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, 3rd Ed. Revised, (DSM-III-R). A.P.A. Washington, D.C., 1987 (trad. franç. sous la direction de J. D. GUELFI, Masson, Paris, 1989).

CARROT G., LANG F., ESTOUR B., PELLET J., GAUTHEY C., WAGON C. — Étude de l'EAT – Échelle d'auto-évaluation de l'anorexie mentale – dans une population témoin et dans une population d'anorectiques. *Ann. Médico. Psychol., 145*, 258-263, 1987.

COOPER Z., COOPER P. J., FAIRBURN C. C. — The Specificity of the Eating Disorder Inventory. *Brit. J. Clin. Psychol., 24*, 129-130, 1985.

CROWTHER J. H., LILLY R. S., CRAWFORD P. A., SHEPHERD K. L. — The Stability of the Eating Disorder Inventory. *Int. J. Eat. Dis., 12*, 97-101, 1992.

FEIGHNER J. P., ROBINS E., GUZE S. B., WOODRUFF R. A. Jr, WINOKUR G., MUNOZ R. — Diagnostic Criteria for Use in Psychiatric Research. *Arch. Gen. Psychiatr., 26*, 57-63, 1972.

GARNER D. M. — *Eating Disorder Inventory-2, Professional Manual, Psychological Assessment Resources Inc.* Odessa, Florida, USA, 48, 1991.

GARNER D. M., GARFINKEL P. E. — The Eating Attitudes Test : an Index of the Symptoms of Anorexia Nervosa. *Psychol. Med., 9*, 273-279, 1979.

GARNER D. M., OLMSTED M. P. — *The Eating Disorder Inventory Manual, Psychological Assessment Resources*. Odessa, Florida, USA, 1984.

GARNER D. M., OLMSTED M. P., DAVIS R., ROCKERT W., GOLDBLOOM D., EAGLE M. — The Association between Bulimic Symptoms and Reported Psychopathology. *Int. J. Eat. Dis., 9*, 1-15, 1990.

GARNER D. M., OLMSTED M. P., POLIVY J. — Development and Validation of a Multidimensional Eating Disorder Inventory for Anorexia Nervosa and Bulimia. *Int. J. Eat. Dis., 2*, 15-34, 1983.

GARNER D. M., OLMSTED M. P., BOHR Y., GARFINKEL P. E. — The Eating Attitudes Test : Psychometric Features and Clinical Correlates. *Psychol. Med. 12*, 871-878, 1982.

HURLEY J. B., PALMER R. L., STRETCH D. — The Specificity of the Eating Disorder Inventory : a Reappraisal. *Int. J. Eat. Dis., 9*, 419-424, 1990.

NORRING C., SOHLBERG — Eating Disorder Inventory in Sweden : Description, Cross-cultural Comparison and Clinical Utility. *Acta Psychiat. Scand., 78*, 567-575, 1988.

RACITI M. C., NORCROSS J. C. — The EAT and EDI : Screening, Interrelationships, and Psychometrics. *Int. J. Eat. Dis., 6*, 579-586, 1987.

ROSEN J. C., SILBERG N. T., GROSS J. — Eating Attitudes Test and Eating Disorder Inventory : Norms for Adolescent Girls and Boys. *J. Consult. Clin. Psychol., 56*, 305-308, 1988.

SHORE R. A., PORTER J. E. — Normative and Reliability Data for 11 to 18 Year Olds on the Eating Disorder Inventory. *Int. J. Eat. Dis., 9*, 201-207, 1990.

SIMON Y. F., WERY C., DIVAC S., CRIQUILLION-DOUBLET S., SAMUEL-LAJEUNESSE B. — LENTCA : Logiciel d'évaluation nosographique pour les troubles des conduites alimentaires. *Ann. Méd. Psychol., 147*, 756-761, 1989.

SZEKELY B. C., RAFFELD P. C., SNODGRASS G. — Anonymity, Sex, and Weight-Preoccupation as Variables on the Eating Disorder Inventory with Normal College Students. *Psychol. Rep., 65*, 795-800, 1989.

WELCH G., HALL A., NORRING C. — The Factor Structure of the Eating Disorder Inventory in a Patient Setting. *Int. J. Eat. Dis., 9*, 79-85, 1990.

WELCH G., HALL A., RENNER R. — Patient Subgrouping in Anorexia Nervosa Using Psychologically-based Classification. *Int. J. Eat. Dis., 9*, 311-322, 1990.

WELCH G., HALL A., WAKLEY F. H. — The Factor Structure of the Eating Disorder Inventory. *J. Clin. Psychol., 44*, 51-56, 1988.

WILLIAMS R. L., SCHAELER C. A., SHISSLAK C. M., GRONWALDT V. H., COMERCI G. D. — Eating Attitudes and Behaviors in Adolescent Women : Discrimination of Normals, Dieters, and Suspected Bulimics Using the Eating Attitudes Test and Eating Disorder Inventory. *Int. J. Eat. Dis., 5*, 879-894, 1986.

25

LIMITES DE L'OPÉRATIONNALISATION

par B. FALISSARD et J. D. GUELFI

Si cet ouvrage s'intéresse, en quelque sorte, à diverses tentatives visant à opérationnaliser la psychiatrie, il est licite de s'interroger sur les limites d'une telle opérationnalisation. En d'autres termes, la psychiatrie s'intéressant à l'homme dans ce qu'il a de plus complexe et informel : son fonctionnement mental, comment peut-on espérer quantifier, numériser ce dernier sans passer à côté de ce qui en fait l'essentiel?

Bien entendu, un tel scepticisme vis-à-vis de l'opérationnalisation de la psychopathologie est compréhensible. Mais il faut avouer qu'il n'existe aucun argument formel susceptible de venir supporter cet *a priori* «naturel». Aussi, plutôt que de se lancer directement dans ce débat, il paraît plus profitable de s'intéresser aux résultats acquis dans des disciplines accomplies. C'est l'objet du présent chapitre qui, précisément, s'intéresse aux limitations conceptuelles insurmontables qui ont pu être révélées dans diverses sciences, qu'elles soient déductives, expérimentales ou humaines.

Nous nous pencherons plus en détail sur trois résultats fondamentaux découverts au cours du XXe siècle : le théorème de Gödel, le paradoxe d'Einstein, Podolski et Rosel (EPR) et le théorème d'Arrow. Ceux-ci se rattachent à trois disciplines bien différentes : la logique mathématique, la physique des particules, et l'économie. Selon le théorème de Gödel, dit théorème d'incomplétude, dans tout système formalisé suffisamment complexe pour pouvoir décrire au moins l'ensemble des entiers, il existe des propriétés vraies que l'on ne pourra jamais démontrer. Le paradoxe EPR (ou plutôt les protocoles expérimentaux qui en ont découlé) aboutit peu ou prou à la démonstration du fait que le caractère probabiliste de la mécanique quantique ne provient pas de notre ignorance mais semble essentiel et insurmontable. Enfin, le théorème d'Arrow montre que lorsque des individus classent chacun une même liste d'objets, aucune technique ne peut résumer de façon satisfaisante l'ensemble de leurs choix en un choix collectif unique.

Ces trois exemples correspondent tous d'une certaine façon à des limitations de l'opérationnalisation.

LE THÉORÈME DE GÖDEL

Si la plupart des sciences expérimentales ont la caractéristique de valider leurs théorèmes lorsque ceux-ci sont en accord avec l'expérience, la logique est, elle, une

science déductive. Dans une telle situation, on considère certaines propositions comme admises sans démonstration (on les dénomme axiomes ou postulats). Les théorèmes sont ensuite dérivés *via* des principes logiques. Un exemple bien connu de science déductive est la géométrie classique, dont les axiomes ont été formulés par Euclide.

C'est vers la fin du XIXe siècle que l'arithmétique a été axiomatisée de manière à lui donner les caractéristiques d'une science déductive. Cette étape a nécessité un travail de formalisation que l'on peut résumer de la façon suivante. On définit un vocabulaire qui permet de construire les formules, les axiomes, qui seront les fondations de la construction; puis des règles de déduction permettront de construire des théorèmes.

Plus précisément :

Le vocabulaire

Il comprend un ensemble non fini de constantes représentant :

– les signes logiques bien connus (quantificateurs, séparateurs ou opérateurs) : « $\neg$ » (négation), « $\wedge$ » (et), « v » (ou), « $\supset$ » (implique), « $\exists$ » (il existe), « $\forall$ » (quel que soit), « (» (parenthèse à gauche), «) » (parenthèse à droite), « , » (virgule);

– « 0 » (le chiffre pour le nombre zéro), et le signe « s » désignant le successeur immédiat d'un nombre (« s0 » pour le successeur de 0, « ss0 » pour le successeur de s0,...);

– un ensemble de variables « x, y, z... » représentant des nombres (ainsi, x peut représenter 0, ss0 ou sssss0);

– un ensemble de variables « p, q, r... » représentant des propositions (p peut représenter « $(\exists x)(x = sy)$ », qui se lit « il existe un x tel que x est le successeur immédiat de « y »);

– et enfin un ensemble de variables « P, Q, R » représentant des relations (dites prédicats). (Q(x) peut représenter « x est un entier » ou encore P(x, y) « x est plus grand que y »).

Les axiomes

Les premiers sont :

$(pvp) \supset p$

$q \supset (pvq)$

$(qvp) \supset (pvq)$

$(p(qvr) \supset (q(pvr)$

$(qvr) \supset ((pvq) \supset (pvr))$

$R(y) \supset (\exists x) R(x)$

Q(0) (ou Q(x) représente « x est un entier »)

$(Q(x)) \supset (Q(sx))$, etc.

Les règles de déduction : substitution et détachement

À partir d'elles, il est possible de formuler de nouvelles propositions ou théorèmes. La première règle autorise la substitution de toute proposition dans toute proposition vraie, par exemple, si (pvp) ⊃ p alors ((pvq) ⊃ (pvq)) ⊃ (pvq) (substitution de pvq à p). La deuxième règle indique que si l'on a p et p ⊃ q, alors on peut en déduire q.

Cette axiomatisation de l'arithmétique connaît pourtant des limites que le théorème de Gödel a pu mettre à jour. Il conduit en effet à montrer qu'en axiomatisant ainsi l'arithmétique, on aboutit inéluctablement à l'impossibilité de démontrer formellement certaines propriétés pourtant vraies. Autrement dit, toute axiomatisation de l'arithmétique est « incomplète ».

La démonstration de Gödel est remarquablement astucieuse, qui consiste en un premier temps à assigner un nombre entier unique à chaque élément du vocabulaire (signes ci-dessus), puis à chaque proposition (succession de tels signes) et à chaque démonstration (succession de propositions). Dans un second temps, l'auteur exhibe un entier n particulier ainsi qu'une proposition p définie par la phrase :

p = « la proposition assignée au nombre n n'est pas démontrable ».

Pour cette valeur précise de n, cette proposition vérifie deux surprenantes propriétés :

– l'expression formelle de p est assignée au nombre n ;
– p est indécidable (on ne peut démontrer formellement ni p ni ¬p).

Maintenant, puisque :

p est indécidable (1)

et que

p est assignée au nombre n,

alors

la proposition assignée au nombre n est indécidable

donc

« la proposition assignée au nombre n n'est pas démontrable » est vraie

donc

p est vraie (*cf.* définition de p ci-dessus) (2)

et pour finir, d'après (1) et (2)

p est vraie et indécidable (CQFD).

Ce théorème nous évoque les remarques suivantes :

Au sein d'une science déductive formalisée comme les mathématiques, il est fondamental de séparer ce qui est métamathématique (à savoir le discours portant sur les propositions mathématiques), des mathématiques proprement dites (signes, axiomes, règles déductives ainsi que les théorèmes qui en découlent). De nombreux paradoxes jouent sur la confusion des niveaux métalogique/logique, par exemple le paradoxe bien connu du menteur :

La proposition écrite dans ce cadre est fausse.

La question est de savoir si la proposition écrite ci-dessus est vraie ou fausse. Si elle est vraie, alors la proposition « la proposition écrite dans le cadre est fausse » est vraie ; or cette proposition signifie justement qu'elle est fausse, donc cela n'est pas possible. Maintenant si cette proposition est fausse, alors la proposition écrite dans le cadre est fausse, donc la proposition « la proposition écrite dans le cadre est fausse » est vraie. Or, par hypothèse, cette proposition était fausse... D'où paradoxe. L'écueil vient ici de l'utilisation d'une proposition qui s'intéresse à son propre sens. Ceci s'avère possible quand on utilise le langage courant car celui-ci sert aussi de métalangage : il sert à se décrire lui-même.

Le point clé de la démonstration de Gödel est l'assignation d'un nombre à chaque proposition et démonstration. *Celle-ci permet d'arithmétiser les métamathématiques.* Chaque proposition étant un nombre, les assertions portant sur les propositions se transforment en assertions sur des nombres, donc à de l'arithmétique... Ce procédé permet d'utiliser en toute bonne foi les paradoxes tels que celui présenté ci-dessus.

La seconde remarque concerne une proposition p, à la fois vraie et indécidable. Ce qui est vrai n'est-il pas justement ce qui est démontré, donc ce qui est décidable?... Pas véritablement. Le vrai est, en effet, une notion sémantique qui n'appartient pas aux mathématiques mais aux métamathématiques. Les mathématiques se contentent de trouver des conséquences logiques des axiomes. Si ces axiomes sont déclarés vrais (déclaration métamathématique), alors les conséquences logiques de ces axiomes sont aussi vraies. Rien ne prouve cependant que tout ce qui est vrai est une conséquence logique des axiomes. L'axiomatisation peut être considérée comme une machine à faire la vérité. Le théorème de Gödel montre que cette machine n'est pas omnipotente, et qu'elle n'épuise pas toute la vérité.

LE PARADOXE EPR

La physique des particules s'intéresse à l'étude des particules élémentaires telles que les atomes, électrons, neutrons, protons ou photons. Ces constituants essentiels de la matière ont longtemps été représentés sous la forme de petits « grains » ou corpuscules » plus ou moins sphériques. Les lois régissant leur interactions étaient les lois habituelles de la mécanique classique : conservation de l'énergie et conservation de l'impulsion. La physique des particules s'apparentait *grosso modo* à la physique des boules de billard...

Vers la fin du XIXe siècle, une série d'expériences est venue remettre en question une telle représentation. Il s'agit notamment de l'expérience dite des fentes de Young ou « d'interférences de photons ». Une source émet des photons vers un écran percé de deux fentes verticales. Derrière lui se trouve un deuxième écran dont le rôle est de visualiser l'intensité de la lumière. Si l'on obstrue une des deux fentes, la lumière apparaît sur le second écran dans une zone en regard de la fente non obstruée. Si les deux orifices sont ouverts simultanément, plutôt que d'observer un résultat identique en regard de chaque fente, l'intensité lumineuse sur le deuxième écran varie périodiquement de telle sorte qu'apparaissent des bandes lumineuses verticales régulièrement espacées. Cette expérience (puis de nombreuses autres) a conduit à conclure qu'il est impossible de considérer qu'une particule passe soit par la fente de gauche soit par la fente de droite mais qu'il faut plutôt considérer qu'elle passe à la fois par les deux fentes produisant ainsi un phénomène d'interférence.

À partir de ce résultat, les physiciens ont proposé une nouvelle description des particules élémentaires. Ces dernières ne sont plus représentées par des petites « boules » mais plutôt par leur probabilité d'apparaître à tel ou tel endroit, quand une mesure vient à être effectuée. Une particule est, d'une certaine façon, diluée dans l'espace, et ce n'est que lorsqu'on essaie de connaître sa position (ou sa vitesse) au moyen d'un dispositif expérimental, qu'elle se rassemble avec une certaine probabilité en un point donné de l'espace et que le paramètre mesuré prend un sens. Un tel point de vue, peu intuitif, a été combattu par plusieurs grands noms de la physique. C'est ainsi qu'en 1935 les trois physiciens Einstein, Podolski et Rosel ont proposé un protocole expérimental imaginaire qui tendait à prouver que si le formalisme de la mécanique quantique permettait en pratique de faire des calculs conformes à la réalité expérimentale, cette théorie était fondamentalement incomplète. Plus précisément, le caractère probabiliste de celle-ci ne serait pas essentiel mais refléterait plutôt l'ignorance des physiciens. Certaines variables gouvernant la réalité sont encore insoupçonnées, non maîtrisées dans les expériences, rendant cette même réalité aléatoire.

Le protocole expérimental imaginé par EPR est le suivant : 2 particules identiques sont placées dans le vide, dans un état où la somme de leurs vitesses est nulle (elles partent dans deux directions opposées). Au bout d'un certain temps, ces deux particules sont très éloignées et le fait de toucher à l'une ne peut plus *a priori* influencer l'autre. Dans une telle situation, on mesure la vitesse de la première particule : la vitesse de la seconde étant exactement opposée à celle de la première, on connaît maintenant la vitesse de la deuxième particule sans avoir aucunement interagi avec elle et naturellement, sans avoir mesuré sa vitesse. Ceci prouve que la vitesse de la deuxième particule est parfaitement déterminée avant toute manœuvre expérimentale. Par conséquent, si le résultat prédit par la mécanique quantique est aléatoire, ce n'est pas parce que la « réalité » l'est mais pour une autre raison, par exemple parce que nous ne maîtrisons pas certaines caractéristiques de la particule ou de son environnement.

En 1964, J. S. Bell prolonge le raisonnement d'EPR et aboutit à un résultat théorique susceptible d'être mis à l'épreuve. Ce résultat, appelé « théorème de Bell » ou encore « inégalités de Bell » a une portée très générale. Il stipule que :

– si deux particules sont créées dans des conditions telles que leurs caractéristiques physiques sont à la fois corrélées et bien déterminées,

– si nous mesurons ces caractéristiques de manière à ce que la mesure effectuée sur une particule n'influence pas la mesure effectuée sur l'autre,

– si enfin ce type d'expérience est renouvelé un grand nombre de fois;

– alors la valeur moyenne, sur l'ensemble des couples de particules, d'une certaine combinaison R des caractéristiques mesurées sur chaque couple de particules doit obligatoirement être comprise entre deux bornes (–2 et 2).

Or, les calculs de mécanique quantique appliqués à un système de deux particules aboutissent, eux, à des valeurs de R en dehors de cette fourchette. Il ne restait donc plus qu'à tester expérimentalement chacune des hypothèses. Ce fut fait dans les années 1970-1980. Les résultats furent clairement une confirmation des calculs de la mécanique quantique, c'est-à-dire une violation des inégalités de Bell. Il est donc nécessaire de réfuter les prémisses de Bell : dans les protocoles expérimentaux décrits ci-dessus, il faut considérer que les caractéristiques physiques des deux particules sont indéterminées ou bien il faut admettre que la mesure effectuée sur la

première particule a un effet sur la deuxième particule. Toute la lumière n'est pas encore faite sur les implications épistémologiques de ces expériences surprenantes.

Retenons pour notre part que la violation des inégalités de Bell conforte sérieusement l'adéquation qu'il existe entre la réalité et le formalisme de la mécanique quantique. Or, par construction, la mécanique quantique exprime des limites quant à la description de la réalité physique, puisqu'elle ne propose qu'un cadre probabiliste. Il s'agit, ici aussi, d'une limitation à l'opérationnalisation.

LE THÉORÈME D'ARROW

La notion de choix collectif est un concept central en économie, intimement liée aux notions de démocratie et de justice. Le théorème d'Arrow peut être explicité dans le cadre particulier d'une procédure de choix électoral. Des citoyens classant individuellement plusieurs candidats, quelle est la méthode qui permet au mieux de synthétiser leurs différents choix afin d'obtenir un classement global de ces mêmes candidats, représentatif de l'ensemble de la société? Nous allons voir qu'en dehors de la situation où il n'y a que deux candidats (où l'on peut appliquer la méthode habituelle du scrutin majoritaire) un tel problème n'a aucune solution « acceptable ».

Une formulation plus précise du théorème d'Arrow est la suivante :

1) Les classements collectifs et individuels doivent être des relations d'ordre (notamment transitives : si A est préféré à B, et B est préféré à C, alors A est préféré à C).

2) Les électeurs ont à classer plus de deux candidats, et toutes les combinaisons de candidats sont autorisées.

3) Le choix collectif n'est pas *systématiquement* le choix d'un seul et même individu (pas de dictature).

4) Le choix collectif n'est pas fixé *a priori* par une règle indépendante de la procédure de vote (société utilisant un ensemble de règles issues de la tradition, code religieux par exemple). Enfin, la procédure de choix collectif doit vérifier deux dernières conditions dites de « cohérence » :

5) Si tous les électeurs décident en bloc de mieux classer un des candidats, la procédure de choix collectif doit fournir un choix de société où ce candidat est ici aussi mieux classé.

6) Imaginons que le scrutin ait lieu, un premier choix collectif est calculé. Si un candidat vient à se retirer subitement (ou décède), en enlevant son nom dans le bulletin de chacun des électeurs, la procédure de choix collectif doit donner un nouveau classement identique au premier choix collectif auquel on a enlevé le nom de ce candidat.

Il est possible de montrer formellement qu'aucune procédure de choix collectif ne peut vérifier conjointement l'ensemble de ces six conditions. Ceci constitue une nouvelle fois une limite à un certain type « d'opérationnalisation ».

Ce théorème peut paraître surprenant. Examinons sur quelques situations concrètes où apparaissent les incohérences.

– *La procédure majoritaire.* Une solution logique serait de considérer qu'un candidat est préférable à un autre si une majorité d'électeurs en a décidé ainsi. Observons la situation suivante : trois candidats X, Y et Z sont en présence, il y a trois

électeurs, notés e1, e2 et e3. Imaginons que e1 vote X, Y, Z; e2 vote Y, Z, X et e3 vote Z, X, Y. Il est ainsi clair que 2/3 des électeurs préfèrent (X à Y) et (Y à Z); le choix collectif devrait donc être X, Y, Z, ce qui est incompatible avec le fait que 2/3 des électeurs préfèrent (Z à X)...

– *Le vote par pondération des classements (Arrow).* Il existe une technique de vote fréquemment utilisée qui consiste à attribuer un certain nombre de points à un candidat quand son nom apparaît en premier dans un vote, à lui attribuer moins de points quand son nom apparaît en second, etc. Au total, les candidats seront classés globalement en fonction du nombre total de points qu'ils auront obtenu. Prenons un exemple où l'on a quatre candidats A, B, C et D et trois électeurs e1, e2 et e3. Un candidat classé premier aura 4 points, un second aura 3 points, un troisième 2 points et un dernier 1 point. Supposons que e1 vote A, B, C, D; e2 vote A, B; C; D et e3 vote C, D, A, B. A a donc 10 points, C 8 points, B 7 points, C 7 points et D 5 points : A est donc élu. Si le candidat B se retire, A a 7 points, C 7 points et D 4 points : A et C sont alors *ex aequo.* La règle numéro 6 n'est pas vérifiée, une telle procédure n'est donc pas acceptable.

DISCUSSION

Les trois points que nous venons d'examiner ont été à l'origine de polémiques sans fin au sein de leurs spécialités d'origine.

Le premier, le théorème de Gödel, est sans nul doute le résultat le plus riche sur le plan épistémologique : il stipule qu'il est impossible de formaliser simplement et de façon satisfaisante une discipline un tant soit peu complexe. De façon plus rigoureuse, il faudrait dire que tout système englobant l'arithmétique des nombres entiers (« ...discipline un tant soit peu complexe... ») et formalisée à partir d'un système de logique finitiste (« ...formalisé simplement... ») est nécessairement incomplet, c'est-à-dire qu'il existe des propositions vraies qui sont indécidables (notion de formalisation satisfaisante).

Nous sommes aux ultimes limites de l'opérationnalisation. Ce résultat majeur ne doit cependant pas nous décourager. En effet, d'une part la psychopathologie est loin de pouvoir revendiquer la qualité de discipline formelle, d'autre part ces limitations théoriques ne sont, pour l'instant, pas gênantes en pratique : aucun mathématicien n'a véritablement été gêné dans ses recherches par l'incomplétude pourtant réelle des mathématiques. Cela est vraisemblablement dû au fait que le caractère incomplet ne se révélerait que dans des situations extrêmes (ce point est cependant discuté).

Le deuxième point est intéressant car il s'applique à une discipline expérimentale, la physique. Dans cette dernière, un cadre théorique en vigueur stipule que certaines caractéristiques de la réalité (position, vitesse de particules notamment) ne peuvent être connues qu'en termes de probabilité. Le paradoxe EPR et la violation des inégalités de Bell ont notamment permis de montrer que ce caractère probabiliste n'est vraisemblablement pas dû à notre méconnaissance de la réalité (existence de variables cachées) mais est en quelque sorte inhérent à cette dernière. Ici, ce n'est donc pas le système de formalisation qui montre ses limites mais plutôt nos possibilités d'appréhender la réalité matérielle. L'étude du fonctionnement mental normal et pathologique n'en est certes pas à fournir des protocoles expérimentaux d'une telle envergure, il n'en demeure pas moins qu'une telle situation est riche

d'enseignements. Si l'étude expérimentale des corps solides inanimés aboutit à de telles limitations, il est (trop) tentant de penser qu'un cadre théorique rigoureux susceptible de décrire le fonctionnement mental pourrait être soumis à des limitations bien plus draconiennes...

Le dernier point de limitation formelle ou de réalité inaccessible concerne plutôt une démonstration de non-existence. L'homme (l'économiste) appelle de ses vœux une technique qui permettrait de refléter fidèlement à l'échelon d'une société les *desiderata* de chacun de ses individus; ce vœu est sans espoir, cette technique n'existe pas. De la même façon, le mouvement perpétuel ou le moteur à eau sont des vues de l'esprit et ne peuvent exister. Nous devons ainsi nous rappeler que toutes les techniques et tous les outils ne sont pas réalisables, même si l'on écarte les éventuelles difficultés pratiques nécessaires pour les mettre en œuvre.

Que penser, maintenant, d'éventuelles limitations à la formalisation de la psychopathologie?

En premier lieu, on peut s'interroger sur la réalité actuelle d'une telle opérationnalisation. En effet, quel que soit le paradigme retenu, nous sommes encore bien loin d'une formalisation de la pathologie mentale. Les modèles psychodynamiques ne revendiquent pas un caractère formel; les cognitivistes n'obtiennent de modèle satisfaisant que dans la description de quelques domaines ponctuels; enfin, les prétendus systèmes sémiologiques athéoriques font parfois penser à un miroir aux alouettes : ils se trouvent, en effet, confrontés aux limitations de la langue qu'ils utilisent, il est même fort possible que l'inconsistance d'un langage naturel soit finalement rhédibitoire.

Enfin, si tant est que cette opérationnalisation soit un jour réalisée, il demeurera toujours quelques bastions inaccessibles, du fait même des limites inhérentes à tout système formel. Voilà qui nous invite à la plus grande modestie...

BIBLIOGRAPHIE

NAGEL E., NEWMAN J. R, GÖDEL K., GIRARD J. Y. — *Le théorème de Gödel.* Seuil, Sources du savoir, Paris, 1989.

LALOË F. — Les surprenantes prédictions de la mécanique quantique. *La Recherche, 182*, 1358-1368, 1987.

D'ESPAGNAT B. — *Une incertaine réalité.* Gauthier-Villars, Paris, 1985.

ARROW K. J. — *Choix collectifs et préférences individuelles.* Calmann-Lévy, Perspectives de l'économie, Paris, 1974.

INDEX ALPHABÉTIQUE

C

D

E

F

G

M

N

O

P

Q

T

MASSON Éditeur
120, boulevard Saint-Germain
75280 Paris Cedex 06
Dépôt légal : septembre 1995

SNEL S.A.
Rue Saint-Vincent 12 - 4020 Liège
septembre 1995